Evaluación y optimización de los programas educativos

M. José Sáez Brezmes

P01/72028/00668



Índice

Introducción

La evaluación de programas es el área más extensa de la evaluación, y la cual, conscientemente, más trabajo teórico ha generado sobre su conceptualización. Tiene una larga historia práctica y fue reconocida como especialidad en Estados Unidos durante la década de los sesenta, aunque antes se habían llevado a cabo algunas evaluaciones en el mundo de la educación que consiguieron una gran relevancia. Por ejemplo, la que hizo Rice en 1897, sobre el estudio de la ortografía de la "b" y la "v" en Filadelfia, y las que realizó R. Tyler entre 1930 y 1940, que lo llevaron a formular las primeras conceptualizaciones y a empezar a utilizar el término.

Los programas están indisolublemente relacionados con el contexto en que se desarrollan; de hecho, la evaluación se limita a observar sus efectos en el tiempo y en un lugar, con un personal determinado, dentro de los límites de una organización determinada y en las circunstancias específicas de un momento histórico concreto, de manera que a veces resulta complicado saber hasta qué punto los resultados observados pueden generalizarse a otras situaciones.

Si la evaluación es hoy día en muchos países una práctica extendida, se debe sobre todo al hecho de que, como explicaba B. MacDonald, tiene como ventaja, ante las investigaciones de tipo aplicado, de que sus informes deben responder a las preguntas y los problemas sobre el objeto evaluado, un programa, por ejemplo, en un tiempo relativamente corto, tratándolos de forma específica y directa, e incluyendo formulaciones concretas de la problemática, a la vez que en muchos casos recomendaciones, más que conclusiones del propio estudio.

Así pues, trataremos la problemática actual de la evaluación de programas educativos, teniendo en cuenta dos factores importantes:

1. El conocimiento que se presenta se ha generado a partir de la práctica de la evaluación externa.

2. La mayoría de la práctica de evaluación y su teorización es el producto del debate desarrollado por muchos autores en Estados Unidos, sobre todo.

En nuestro país hay todavía una experiencia reducida en el campo de la evaluación de programas y sólo desde hace poco han empezado, en algunas universidades, cursos o másteres que se encargan de la formación de evaluadores, pero hay indicios claros en nuestro sistema social de que la evaluación tiene ya un espacio importante para ocupar en el sector educativo. Por ello, éste es un momento interesante para estudiarla, porque ofrece la posibilidad de contribuir a sus primeros pasos. En Europa la Asociación de Evaluación se creó en 1994.

Objetivos

Los principales objetivos que hay que alcanzar en este módulo didáctico son los siguientes:

1. Conocer cuándo y cómo surge la evaluación de programas.

2. Describir la evolución del concepto de programa.

3. Conocer las características de los enfoques más importantes de la evaluación de programas.

4. Explicar los factores que han hecho evolucionar los principales modelos.

5. Enumerar las preguntas básicas que orientan los diseños.

6. Describir las fases de las que consta la evaluación de programas.

7. Conocer las denominaciones y características de las audiencias.

8. Explicar los criterios que orientan la realización de un informe.

9. Establecer las diferencias que hay entre evaluación interna, evaluación externa y autoevaluación.

10. Explicar los principios éticos que rigen la actuación de los evaluadores como profesionales.

Objetivos complementarios:

- Distinguir entre evaluación e investigación.
- Estudiar otros enfoques de evaluación.
- Hacer un diseño para la evaluación de un programa.
- Estudiar la evolución del concepto de negociación.
- Desarrollar criterios para la discusión y presentación de un informe.

1. La evaluación de programas educativos

1.1. ¿Qué se entiende por programa educativo?

R. Stake considera **programa** toda actividad organizada que se prolonga en el tiempo para conseguir unos objetivos, que cuenta con un sistema de gestión y financiación y que va dirigida a un grupo de individuos, y despierta el interés de muchos otros (*stakeholders*).

Sin embargo, muchas veces los programas no son entidades perfectamente definidas por adelantado, sino que se les denomina de esta forma porque están encaminados a la consecución de unos objetivos que pretenden resolver un problema o atender unas necesidades de política social. Por este motivo, con el término *programa* podemos encontrar significados y dimensiones diferentes; así, un programa puede ser tan amplio como la formación permanente de los profesores de una región o un país, o tan limitado como las actividades de los profesores que atienden a los alumnos de diversificación, planificadas por el departamento de orientación de un centro escolar.

Los **programas sociales** son empresas extremamente complejas, y sus evaluadores muestran cierto recelo hacia quienes se encargan de hacer evaluaciones para otros sectores que cuentan con parámetros físicos que se pueden medir y, sobre todo, reproducir, como la medicina, cuando el objetivo es estimar los efectos de un medicamento. En este caso, el mismo estímulo provoca una y otra vez efectos similares, de manera que otros investigadores en estas u otras circunstancias pueden estudiar sus consecuencias y trabajan con la seguridad de que siempre están analizando los efectos de una misma cosa, un producto.

Los programas de acción social...

... están divididos en tres grupos:

1. Programas educativos.

2. Programas de salud.

3. Programas que se refieren a la aplicación de las leyes.

Los primeros son una parte importante de la industria de los servicios humanos; su desarrollo ha tenido un gran crecimiento en los últimos veinte años, y está vinculado a la idea de las reformas, las innovaciones curriculares y el cambio en la educación.

Los programas sociales, desgraciadamente, no son tan específicos, porque abarcan una amplia gama de componentes, estilos, personas y procedimientos, como por ejemplo un nuevo enfoque curricular para la enseñanza de las ciencias en algún nivel educativo concreto. De hecho, al evaluador a menudo le resulta difícil describir realmente cuál es el contenido o qué está pasando en el "programa", al menos al comienzo de su trabajo.

A lo largo de los casi cuarenta años de la historia de la evaluación, algunos evaluadores han considerado los programas como una caja negra, cuyos contenidos no eran el principal objetivo, sino descubrir sus efectos. Esta conceptualización del programa presenta algunos problemas, porque si el evaluador no lo conoce bien, no sabrá cuál es la problemática más importante que tiene, y por lo tanto no podrá formular en su investigación las preguntas más

adecuadas que afectan a los aspectos sustanciales del programa. También ha llevado, a veces, a buscar unos efectos faltos de realismo, porque la propaganda que se ha hecho del mismo hace esperar resultados de mayor envergadura o porque, a pesar de lo que los promotores anuncian verbalmente, los recursos se han distribuido en la práctica de manera diferente. ❗



A lo largo de casi cuarenta años, algunos evaluadores han considerado los programas como una caja negra.

Si el evaluador no sabe en qué consiste el programa, tampoco sabrá a qué atribuir los resultados que observa, ya que es importante no olvidar que en la lógica de la evaluación se encuentra el hecho de ofrecer una información que ayude a tomar decisiones. Para ello es preciso conocer cuáles son las actividades que produjeron determinados resultados y cuáles las que no lo hicieron, o aquello que no sirvió para casi nada. Por este motivo, el evaluador debe tener datos de la naturaleza del programa tal y como existió, y no tan sólo según lo describieron los financiadores, porque si no es así, no habrá ninguna información en que pueda basarse algún tipo de decisión. Imaginemos que se evalúa un programa que tiene un gran éxito, y los evaluadores quieren recomendar su extensión a otros contextos: ¿qué es lo que se podrá recomendar que se implante en otros lugares?

De este modo, el trabajo del evaluador es descubrir la realidad del programa y descartar sus versiones idílicas, pero a veces los programas que se evalúan son muy vagos o se definen a medida que se desarrollan, o son muy nuevos en cuanto a estructura y organización, por lo que es necesario que el evaluador recopile información para describir con realismo lo que está pasando. Para hacerlo tiene que establecer una serie de procedimientos que le permitan conocer el programa en acción, y reunir información de manera sistemática.

Los procedimientos...

... para conocer el programa en acción pueden ser sencillos, como el estudio y la revisión de los documentos publicados por el mismo durante su desarrollo, o una serie de entrevistas al director y a algunos profesores clave en el caso de un programa educativo, o asistir a reuniones ordinarias, etc.

Cuanto mayor sea el programa en lo que respecta al número de individuos que trabajan en él, los espacios geográficos en que se aplica o el tiempo que lleva en práctica, más necesario será establecer un muestreo sistemático que permita caracterizarlo. En la mayor parte de los programas es suficiente con unas cuantas categorías descriptivas que sean adecuadas al caso para asimilar la esencia del programa, como por ejemplo:

En las evaluaciones...

... realizadas en los últimos años ha habido un énfasis creciente para definir esmeradamente los programas. A ello han contribuido de manera notable los evaluadores cualitativos, que han convertido este trabajo en un objetivo inicial, pero esencial, de sus evaluaciones, lo cual ha hecho que los evaluadores que trabajan con métodos cuantitativos pongan más interés en este tipo de evaluación

1) el tipo de servicio que el programa ofrece,

2) dónde se pone el énfasis conceptual,

3) la clase de personal que trabaja en él,

4) el marco en que se desarrolla,

5) la organización que lo patrocina.

Ya hace un cierto tiempo, cuando los **enfoques experimentales** eran los que se aceptaban mayoritariamente, algunos autores (Chen y Rossi, 1984) argumentaron en contra de la idea de la "caja negra", como conceptualización para la definición del programa, que dominó durante un par de décadas la literatura de la evaluación, y sugirieron que la definición debería realizarse de acuerdo con la teoría de las ciencias sociales. McClintock (1987) ha publicado un estudio en el cual resume las conceptualizaciones heurísticas de los programas que incluyen metáforas, modelos causales y mapas conceptuales.

Hay autores que consideran los programas como entidades sumamente complejas producidas por una gran cantidad de factores que interactúan entre ellos, y que producen a la vez una importante variedad de resultados. Determinar las relaciones contingentes con el programa y sus resultados no es una cosa fácil; un programa no es una máquina que da un producto muy definido. Pensemos en un ejemplo sencillo: el curso de un profesor que utiliza el mismo temario, los mismos libros y materiales, la misma estrategia de enseñanza y aprendizaje y que da clase a dos grupos de estudiantes, que pueden ser, sin embargo, muy diferentes por la dinámica que se establece en cada uno de ellos, y que aprenden cosas distintas.

E.R. House (1993) argumenta que el método y el enfoque utilizados para hacer la evaluación no pueden sustituir la naturaleza o lo que es sustancial del programa. Según él, lo que se ha constatado a lo largo de la historia de la evaluación es que el mundo social es mucho más complejo de lo que podía pensarse cuando empezaron a evaluarse los programas, por lo que una fuente importante de los errores que los evaluadores cometen se debe a la existencia y el uso de una versión estandarizada y simplificada de la teoría de la causalidad. Según cree, se debe abandonar la teoría de la regularidad de Hume para conceptualizar mejor los programas.

Esto no tiene que hacer pensar que los métodos experimentales o prácticamente experimentales no son los más adecuados para evaluar determinados programas, sino que los resultados deben interpretarse con más cautela y que es necesario que los evaluadores sean más modestos en sus conclusiones, porque estos métodos, y en general todos los que se utilizan, no pueden establecer unas relaciones causa-efecto tan precisas. Por lo tanto, la forma en que se define el programa afecta de manera sustancial a la evaluación, y por ello nace la necesidad de convertir el programa en objeto de estudio.

Durante un tiempo, los programas también se consideraron como entidades invariables y su variabilidad como una fuente de error. Cook y Campbell, en 1979, defendían que los programas tuvieran un tratamiento y una implementación lo más estandarizados que fuera posible en sus múltiples aplicaciones.

Sin embargo, esto no deja de ser simplemente un deseo loable para mantener el rigor del proceso de investigación. Las evaluaciones llevadas a cabo consta-

tan la tendencia que tienen los programas a cambiar, algunas veces de manera sutil, otras de forma repentina y drástica, a causa de múltiples factores, como la modificación de la asignación presupuestaria, que permite incrementar o reducir recursos o actividades, el personal y quienes lo sustituyen, que mantienen puntos de vista diferentes y hacen su trabajo con distinta aptitud, los políticos de que dependen, lo cual en ocasiones hace que se modifiquen considerablemente los principios que orientan los programas, etc.

En general, se puede decir...

... que cuanto más largo y complejo es un programa, más posibilidades tiene de cambio. Esto hace que el evaluador deba estar en permanente contacto con los miembros del programa para saber qué cambios se producen y qué efectos tienen sobre las categorías descriptivas que había utilizado.

La actitud que el evaluador debe tomar con respecto a la definición del programa ha sido motivo de importantes polémicas durante la década de los sesenta y los setenta, que han concluido con dos ideas, aceptadas hoy día por toda la comunidad evaluadora. Según estas ideas, la tarea del evaluador debe:

1) Perfilar un modelo dinámico del programa que permita asignar categorías y su conceptualización en función de sus cambios.

2) Estudiar no solamente los resultados, sino el propio proceso de desarrollo del programa y cómo se generan estos resultados.

Lectura recomendada

Para ampliar esta información podéis consultar:

C.H. Weiss (1975). *La investigación evaluativa: métodos para determinar la eficiencia de los programas en acción* (pág. 117). México: Trillas.

Cuando el objeto que se somete a evaluación es un **programa de un organismo público**, como la mayoría de los programas educativos en nuestro país, el evaluador debe tener en cuenta que su evaluación tiene que satisfacer las normas de los procedimientos de decisión previstas en nombre del interés público, pues hay que preguntarse lo siguiente: ¿para quién se hace la evaluación?, ¿para el gobierno?, ¿para el público?, ¿para los beneficiarios del programa?, ¿para quienes lo gestionan y dirigen?

La evaluación, en este caso, forma parte de una situación pública; ya no se trata sólo de un asunto interpersonal, sino colectivo, lo cual hace que se deba tener en cuenta la comunidad de intereses colectivos pertinentes. De hecho, éste es el problema fundamental del evaluador moderno: su evaluación no sólo tiene que ser veraz y creíble, sino que también debe ser justa, porque la evaluación es una empresa que tiene lugar en un contexto político concreto.

C.H. Weiss, que ha estudiado el uso de la evaluación en departamentos ministeriales, plantea desde sus primeros escritos que la evaluación de programas es una tarea política por sí misma, lo cual atribuye principalmente a tres factores:

1) Los programas son creaciones de las decisiones políticas, que se mantienen o no, por presión de determinadas fuerzas políticas.

2) Dado que la evaluación debe informar de la toma de decisiones, tiene un papel en un escenario claramente político.

"La evaluación asiste a...

... los que toman las decisiones a elegir sobre cómo deben irse desarrollando los acontecimientos. Datos precisos y no sesgados sobre las consecuencias que tienen las actividades de los programas deberían mejorar la toma de decisiones, porque la evaluación es una empresa racional que se lleva a cabo en un contexto político." (Weiss, 1975, pág. 37).

3) La evaluación trabaja dando información sobre la legitimidad de los objetivos de un programa y sus estrategias, y sobre la utilidad de éstas para conseguir reformas.

"Las consideraciones...

... políticas influyen a tres niveles, y si el evaluador no las reconoce, sólo le causará disgustos y frustraciones." (Weiss, 1975, pág. 37.)

1.2. ¿Cómo y para qué pueden evaluarse los programas educativos?

Todo es susceptible de ser evaluado, incluida la misma evaluación, pero ahora debemos tratar sobre la evaluación de programas educativos y las políticas que los generan y justifican, cuyo objetivo es el bienestar mediante el conocimiento y la educación de los individuos, las organizaciones y la sociedad. Por ello es importante valorar cómo lo hacen los programas, y cómo pueden hacerlo de la manera más eficaz; por este motivo, también, la necesidad de procesos de retroalimentación *(feedback)* es una presión real, pero tenemos pocos criterios en los que se esté de acuerdo para juzgar la importancia y el valor de las actividades educativas. Porque ¿en qué consiste la mejora de la calidad de la enseñanza o la educación?

La evaluación de programas es un conjunto de métodos, destrezas y sensibilidades para determinar si los servicios prestados son necesarios, si se utilizan, si son lo suficientemente intensos como para suplir la necesidad identificada, si el servicio se da en los términos en que se planificó y si ayuda realmente a la gente que lo necesita dentro de un coste razonable, y sin provocar, además, efectos no deseados.

Cuando se encarga la evaluación de un programa, tal y como hemos comentado en el apartado anterior, es necesario, además de definirlo, saber cuál es la tarea que el evaluador tiene que llevar a cabo en él. Por ello, conviene no confundir la evaluación con la investigación, ni con una valoración de tipo individual:

En el caso de la **investigación**, ésta tiene como objetivo fundamental formular y contestar preguntas de tipo teórico sin preocuparse de las necesidades particulares, concretas e inmediatas de las organizaciones, los *policy-makers* o los directores del programa que piden la evaluación.



La evaluación tiene los siguientes objetivos:

1) Proporcionar información que ayude a los que trabajan en los programas a optimizar su eficacia, tanto en la forma de trabajar como en los resultados.

2) Ayudar también a los que trabajan para la administración y deben tomar decisiones sobre los programas.

3) Contribuir a hacer que los programas sean conocidos y a rendir cuentas al público que los financia.

Si los evaluadores y la evaluación no cumplen una o más de las necesidades expuestas y no proporcionan ninguna información, la evaluación no está cumpliendo su principal objetivo.

B. MacDonald, un evaluador con una gran experiencia, nos hace la siguiente advertencia:

> "... pero nunca cometáis el error de contestar a las preguntas que nadie excepto el evaluador se hace. Lo primero que se debe hacer (en una evaluación) es identificar las preguntas significativas y sólo entonces abordar las cuestiones de tipo técnico que hayan surgido. Limitar la investigación a satisfacer los cánones críticos de la investigación tradicional es arriesgarse a no coincidir con el vocabulario y la terminología de la acción que utilizan los que toman decisiones... El peligro de conceptualizar la evaluación como una rama de la investigación es que los evaluadores queden atrapados en los restrictivos tentáculos de la respectabilidad investigadora."

B. MacDonald (1983). "Evaluación y control de la educación". *Acción Educativa*.

Hay muchos factores que...

... diferencian la evaluación de la investigación, como la universalidad del fenómeno estudiado, la base disciplinaria en que una y otra se fundamentan, cómo vive el evaluador esta experiencia a diferencia del investigador, la autonomía que uno y otro tienen cuando desarrollan su trabajo, etc.

Lectura recomendada

Para ampliar esta información podéis consultar:
N. Norris (1991). *Understanding Educational Evaluation* (pág. 98). Londres: Kogan Page.

La otra actividad profesional que a menudo se confunde con la evaluación de programas es la **valoración individual**. De hecho, hay una larga tradición en la psicología que proviene de información diagnóstica para las organizaciones sobre el tipo de aptitud, intereses, inteligencia, éxitos y resultados de tests de personalidad, etc. de los servicios humanos, es decir, todo lo que una persona necesita para la cualificación de un lugar de trabajo; sin embargo, esta actividad no es parte de la información que el evaluador de un programa tiene que dar.

Veamos la perspectiva histórica de cómo surgió la evaluación de los programas educativos para determinar y comprender cómo y para qué se hace actualmente la evaluación de los mismos.

En 1929, **Ralph Tyler**, que acuñó el término, fue invitado a participar en un programa para mejorar la enseñanza de alumnos que no se habían graduado en la Universidad de Ohio. En este programa empieza a verse la importancia de los objetivos en el diseño curricular. Más tarde, Tyler fue invitado a dirigir la evaluación de un programa piloto sobre la eficacia de la enseñanza secundaria, lo cual le llevó a plantearse el valor de esta enseñanza en lo que respecta a la formación que proporcionaba a los estudiantes para la entrada en la universidad o para acabar en otro tipo de estudios o en el trabajo.

En esta evaluación, Tyler valoró el aprendizaje de los estudiantes, pero remarcó que el término *evaluación* tenía un significado más amplio que el simple hecho de medir, pasar tests o examinar, porque lo que estaba haciendo realmente con su trabajo era determinar hasta qué punto se habían conseguido los objetivos de esta formación y validar las hipótesis con que las instituciones operaban.

Así, Tyler propuso una teoría de la evaluación en la cual ésta aparecía integrada en el proceso de construcción del currículum, marcando las barreras de evaluación y desarrollo del currículum de manera difusa. Para Tyler, la educación

tenía un importante componente de cambio de los modelos de comportamiento de los alumnos y, por ello, el currículum podía construirse mediante la especificación de comportamientos deseables que se establecían como los objetivos que querían alcanzarse. Por este motivo, se debía centrar en buscar hasta qué punto éstos se habían conseguido.

A medida que la demanda de evaluación se incrementó, parte de la idea inicial de Tyler se fue perdiendo, sobre todo la que hacía referencia a la participación del profesorado y su ubicación en el contexto. Después de unos años se reanudó, adoptando la idea de la construcción sistemática del currículum basado en objetivos de comportamiento y aprendizaje, con la ayuda de la tecnología de los tests. De hecho, a principios de siglo parecía que los tests ofrecían un procedimiento científico para medir la capacidad intelectual; la **psicometría** era la ciencia básica de la educación. Algunos autores señalaron de forma muy rápida la potencialidad de la psicometría para contribuir al progreso de la humanidad y llegar a la igualdad de oportunidades. La psicometría prometía también una tecnología para medir los *inputs* y *outputs* educativos y para juzgar la eficacia y eficiencia de los diferentes profesores, escuelas, currículos y programas. Así, la tecnología de los tests proporcionaba la base científica para las comparaciones y la clasificación social.

Los trabajos de...

... Francis Galton (1822-1911) en Inglaterra y Edward Lee Thordinke (1874-1949) en Estados Unidos contribuyeron a extender la tecnología de los tests de medida de la capacidad intelectual.

La National Science Foundation...

... en los años cincuenta, empezó a financiar proyectos curriculares de física, química y biología con fuertes inversiones estatales, lo cual hizo que aumentara el interés por la evaluación, tanto para saber cuáles debían continuar siendo financiadas, como para apoyar su proceso de desarrollo.

Hasta el año 1965, la evaluación era considerada como una actividad relativamente minoritaria de algunos profesores de universidad y de agencias de consultores. La aprobación de la Ley de la enseñanza primaria y secundaria en Estados Unidos cambió el panorama, porque el senador R. Kennedy insistió en que esta ley tenía que incluir una adenda en la que se explicitara la práctica de una evaluación como mandato federal en un programa educativo. Este requisito se extendió rápidamente a otros programas sociales y, posteriormente, a otros países. Desde entonces ha ido creciendo de manera imparable, como lo demuestra el número de las publicaciones existentes, las asociaciones profesionales en Estados Unidos, en Canadá, en Europa, etc., con un ascendente número de miembros.

McLaughlin, que estudió...

... el origen de la evaluación en Estados Unidos, señala que el senador Kennedy, preocupado por el destino de la financiación dedicada a programas educativos para atender a niños discapacitados (*Follow Through*), quiso asegurarse de que el dinero se gastaba en ellos y no en cuestiones de tipo general en los centros escolares. Y lo hizo forzando a las escuelas a pasar cuentas del rendimiento de estos alumnos a sus padres, pensando que ellos, con la información recibida, presionarían a los centros escolares para mejorar la educación de sus hijos. El motivo era, sin duda, una falta de confianza en los métodos tradicionales de enseñanza y en los mismos educadores, y la intención, proporcionar "información científica" que permitiera a los padres reclamar y ejercer una cierta presión para mejorar.



En la foto, Robert Kennedy.

Hasta este momento, la evaluación casi se había reducido al ámbito educativo y consistía en "la evaluación" del aprendizaje mediante los tests, que los llevaban a cabo especialistas en contenidos, psicólogos, sociólogos y personal de la administración.

Desde entonces se han producido grandes cambios en la evaluación de programas, que ha desarrollado una visión más pluralista del concepto de evaluación en lo que respecta al uso de múltiples métodos, criterios, perspectivas, audiencias e intereses, que E. House atribuye al hecho de que la evaluación necesita representar la inmensa heterogeneidad del entramado social de la sociedad americana.

Hay muchas y diferentes intenciones a la hora de evaluar los programas educativos:

a) A veces la evaluación es parte de una actividad de seguimiento de la propia institución que desarrolla e implementa o diseña el programa para saber si éste es eficaz o no, y si no lo es, introducir modificaciones sustanciales en el diseño.

b) Otras veces la evaluación se diseña con el objetivo de identificar las desventajas y los problemas que el programa tiene en acción, con el objetivo de que la información que se reúna sirva para ser utilizada en otros programas cuando se diseñe su financiación y mantenimiento futuros.

c) A veces un programa se experimenta y se lleva a cabo en centros pilotos en diferentes contextos. Así, la evaluación se diseña con la finalidad de que se identifiquen aspectos que no producen los resultados esperados, lo cual sirve a los diseñadores y constructores del programa para modificarlo en un esfuerzo de mejorar su eficacia.

d) Se utiliza también para identificar los efectos diferenciales en distintas poblaciones de estudiantes cuando las diferencias que les afectan, como por ejemplo la clase social, las diferencias étnicas o el nivel adquisitivo de las familias se cree que afectan a la eficacia del programa.

e) Además, algunas instituciones y algunos investigadores la utilizan como parte de sus procedimientos para revisar la relevancia y la validez de los principios en que se basa el programa, porque estos principios provienen de generalizaciones hechas a partir de observaciones sobre una infinidad de prácticas educativas. Y sin embargo, los programas no pueden ser fácilmente generalizados, ya que se modifican de forma consciente o inconsciente cuando se implementan, pues interactúan tanto con los estudiantes como con los otros grupos interesados, ya sean padres, formadores, administradores, etc.

Con frecuencia, la evaluación de...

... programas la solicita el centro principal de una red de centros periféricos ubicados en otras zonas geográficas, como por ejemplo la red de centros de profesores y de recursos, con la finalidad de ayudarlos a determinar y medir el valor, las limitaciones y el coste de los programas que ponen en funcionamiento.

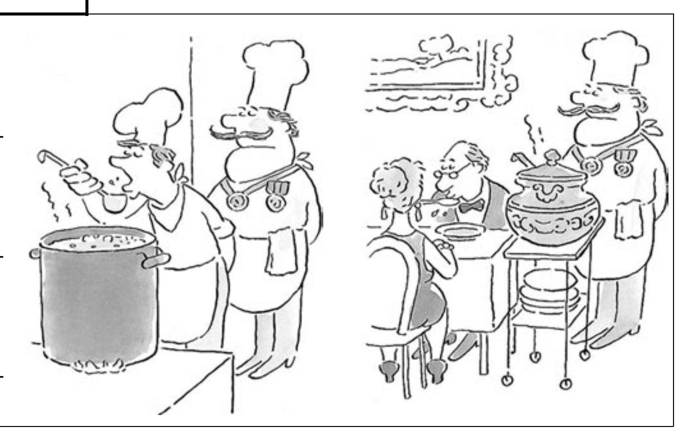
La evaluación de programas puede ser una herramienta muy poderosa para mejorar la eficacia de las organizaciones y optimizar los recursos de que disponen. Por lo general, los programas pueden evaluarse de dos maneras: formativamente y sumativamente, si seguimos la terminología propuesta por M. Scriven en 1967 para ilustrar los dos papeles diferentes de la evaluación:

– La **evaluación formativa** es la que está diseñada y pensada para apoyar el proceso de mejora, y está encargada y llevada a cabo por quienes pueden hacerla.

– La **evaluación sumativa** se centra en el programa como totalidad y en su impacto, sus resultados y sus conclusiones.

B. Stake ejemplarizó

... la diferencia entre las evaluaciones formativa y sumativa al comparar las funciones de un cocinero, que prueba la sopa que preparó antes de servirla a los comensales de su restaurante –evaluación formativa–, y de los comensales cuando la prueban –evaluación de carácter fundamentalmente sumativo.



Lectura complementaria

Resulta interesante seguir el análisis de M. Scriven sobre las diez falacias que la distinción entre evaluación formativa y sumativa ha originado, y que hacen reflexionar sobre la naturaleza de la misma evaluación. Consultad la siguiente obra:
M.W. McLaughlin; D.C. Philips (1991). "Evaluation and Education: a quarter century". NSSE.

Sin embargo, esta distinción ha generado muchos y más detallados estudios desde entonces. ¿Se debe esto a que ha comportado muchas confusiones, como cuándo tiene que hacerse una u otra? ¿Las dos deben incluir recomendaciones?

1.3. ¿En qué momento puede evaluarse un programa educativo?

Examinar la relación entre el desarrollo de un programa educativo en acción y la práctica evaluadora implica analizar la relación entre el objeto y el sujeto de la evaluación. Ésta es una relación compleja porque puede adoptar formas diferentes.

Pueden establecerse al menos tres formas de relación entre el programa en desarrollo y la evaluación:

1) **Relación de tipo causal:** la evaluación se planifica desde el inicio del programa y se establece que los cambios que la evaluación debe adoptar sean los

cambios en la práctica del programa. Es un buen ejemplo de ello la relación orgánica que se establece entre evaluación y currículum en el desarrollo curricular, y que en el proceso se considera una evolución. La implicación es la que hay en los modelos de Tyler y Bloom, que han sido muy utilizados en el desarrollo curricular. Su atributo principal es que la evaluación ofrece líneas maestras de actuación para los que llevan a cabo el desarrollo del currículum.

2) **Relación de tipo fortuito:** la relación que se produce entre las dos se atribuye al azar o al impacto de factores de tipo externo. Así, puede haber interacciones o, incluso, raíces paralelas entre el desarrollo del programa, un currículum y los contextos de las estrategias de la evaluación, que aun actuando por separado o independientemente dan lugar a un proceso interactivo fructífero.

3) **La evaluación y el programa son tareas independientes**, cada una con sus propios intereses, criterios y puntos centrales de atención. En el "enfoque libre de objetivos" defendido por M. Scriven, los evaluadores evitan conocer los objetivos del programa y se centran en estudiar sus efectos reales más que los resultados de lo que se habían propuesto.

De esta manera, la evaluación y el programa trabajan con absoluta independencia y desconexión de quienes desarrollan el programa, aunque este modelo pocas veces se utiliza en el campo educativo.

Sin embargo, aunque el programa esté bien especificado es incluso posible que una evaluación de tipo sumativo no sea recomendable en cualquier momento.

Pueden considerarse tres fases en la vida de un programa: 

1) **Fase en la cual se establecen las condiciones básicas del programa**. En ella es más recomendable una evaluación formativa relacionada con el desarrollo del programa.

2) **Fase**, común a todos los programas, **en la cual se establecen las regulaciones y los procedimientos operativos**, cuando los profesores y el centro escolar en su conjunto se familiarizan con el programa, los objetivos y las estrategias. Se trata de un momento en el cual el programa necesita una gran flexibilidad para adaptarse a las condiciones del contexto y desarrollar las características particulares para cada lugar, por lo que tampoco es recomendable una evaluación sumativa, ya que probablemente coartaría la aparición de ideas innovadoras. Sin embargo, pueden ser adecuadas una evaluación de rendición de cuentas y también las formativas, que ayuden a guiar su adaptación al contexto. Se han observado efectos negativos en evaluaciones sumativas realizadas en los programas en esta fase.

Debate actual

Continúa siendo motivo de estudio y argumentaciones qué tiene que saber un evaluador sobre el área a la cual pertenece el programa para evaluarlo (por ejemplo, qué debe saber sobre educación).

3) En algunos programas hay objetivos de impacto –lo que sería la tercera fase– que no pueden evaluarse hasta que no ha pasado un tiempo prudencial de

funcionamiento, como por ejemplo los que se refieren a salidas profesionales de determinados planes de estudios; en este caso particular, aunque algunos resultados pueden verse con cierta perspectiva relativamente temprana, el impacto de los planes no puede conocerse completamente hasta que el programa lleva funcionando unos cuantos años. Sin embargo, conviene indicar que este tipo de evaluaciones suele ser extraño, porque lo que se acostumbra a pedir a los evaluadores son resultados a corto plazo, y éstos no encuentran financiación con facilidad.

Lectura recomendada

El capítulo del libro que se reseña a continuación puede consultarse en la web.

L.J. Cronbach y otros (1980). *Toward Reform of Program Evaluation: Aims, Methods and Institutional Arrangements* (pág. 111). San Francisco: Jossey-Bass.

2. Modelos y métodos para la evaluación de programas educativos

2.1. Principales modelos y métodos

Tan pronto como empezamos a adentrarnos en el estudio de la evaluación, encontramos los modelos o enfoques de evaluación y observamos cómo se han identificado en muchas ocasiones con métodos de evaluación que hacen clara referencia al momento en que los planteamientos de la evaluación estaban dominados por las formas de investigación.

Según Schuman,...

... "(la evaluación) es principalmente investigación, y por ello debe seguir tanto como sea posible los estándares aceptados de la metodología de la investigación" (Schuman, 1968).

Schuman planteaba también que, adoptando el método científico, un evaluador conseguía unos descubrimientos y unos resultados objetivos, y que la investigación, por lo tanto, era fiable y válida.

La misma definición de modelo es polémica entre los evaluadores, pero podríamos utilizar la que señala su característica principal: ser la representación estructural de la realidad que trata de explicar. La elección del modelo supone la elección de una teoría epistemológica de aprehensión de la realidad.

> Los modelos sirven fundamentalmente para conceptualizar el campo de evaluación y proporcionar un lenguaje común que permite debatir los puntos y temas más conflictivos. Los modelos también establecen los límites de la actuación de los evaluadores, y sus características los hacen más adecuados para determinadas situaciones. Por este motivo, los evaluadores deben conocerlos todos con el fin de utilizar el más adecuado a la situación que se presente.

La mayoría de los evaluadores aplica alguna variante de los modelos más conocidos cuando evalúa un programa. Los tipos básicos son diseños o tipos estructurales ideales, utilizados y recomendados por algunos teóricos de la evaluación más conocidos (en la bibliografía podéis encontrar discusiones y debates sobre sus argumentaciones). Otros autores se refieren a ellos como enfoques, pues son formas genéricas de tratar la evaluación de programas; de este modo, utilizaremos indistintamente uno u otro término cuando hablemos. Seguiremos en esta exposición la secuencia que nos ofrece una explicación histórica de su aparición, que nos permitirá ver cómo responden a preguntas esenciales que la evaluación tenía planteadas en el momento en que cada autor trató la utilización, la descripción y el desarrollo de su modelo.

Consultad el apartado 1 de este módulo didáctico.


Ya vimos los orígenes de la evaluación y la primera conceptualización realizada por R. Tyler, cuyo enfoque, **objetivos de comportamiento**, se propuso en los años cincuenta:

La conceptualización de...

... Tyler es, quizá, la más popular en las evaluaciones de programas educativos, a pesar de las críticas que han recibido el uso y la especificación de objetivos (consultad a Eisner en J. Gimeno, 1982).

1) La evaluación se encarga de definir los resultados y objetivos del programa en forma del comportamiento individual de los estudiantes, de tal manera que

la tarea del evaluador consiste en ver hasta qué punto los estudiantes muestran estos comportamientos una vez que el programa se ha llevado a cabo.

2) El evaluador especifica los objetivos de forma mensurable, los utiliza como la fuente principal de criterios y compara, mediante diseños experimentales o casi experimentales, lo que el programa pretende alcanzar con lo que ha conseguido realmente. La diferencia entre los objetivos y los resultados es lo que da la medida del éxito del programa.

Bloom construyó una taxonomía de los objetivos educativos, con lo cual centró el interés en la creación de la tecnología de cómo se medían estos objetivos y en el desarrollo de tests referidos a criterios con esta finalidad. El desarrollo de la elaboración de tests como una industria continúa dando un apoyo muy importante a este enfoque.

El principal problema de la concepción de Bloom...

... consiste en que cree que la educación es un conjunto de técnicas que llevan a una finalidad preespecificada. Otro problema es la definición misma de los objetivos, que representan unos intereses que la mayoría de las veces son ajenos a los estudiantes. Si se analizan las evaluaciones practicadas, se observa que habitualmente se definen de forma arbitraria.

La inclusión de la evaluación en programas educativos por mandamiento gubernamental generó el enfoque conocido como **análisis de sistemas**, que tiene su principal antecedente en el análisis desarrollado por economistas en el ministerio de defensa americano, y que se ha utilizado para evaluar programas de salud, educación y bienestar social desde 1965.

El modelo se basa en la visión del mundo de la microeconomía aplicada de esta manera al área de los servicios humanos. El análisis coste-beneficio es el más distintivo del enfoque.

La información se traduce en datos numéricos, con los cuales trabaja de forma agregativa más que con sus distribuciones. En la evaluación del proyecto *Follow Through*, prototipo de este enfoque, los resultados se presentaban en una curva normal de equivalencia, y cada estado americano aparecía representado por una sola puntuación. Se asumía la existencia de un acuerdo sobre los objetivos del programa de todos los sectores que participaban en el mismo, por lo que el problema se reducía sobre todo a cómo medir los resultados del programa gubernamental. Utilizaban diseños experimentales en los cuales basaban la relación causa-efecto, y elegían el **grupo control** al azar.

Lectura complementaria

El siguiente libro refleja perfectamente el modelo del cual se habla en el texto:

P.H. Rossi; H.E. Freeman (1993). *Evaluation. A Systematic Approach*. Newbury Park, Londres y Nueva Delhi: Sage.

Rossi y Freeman (1993) plantean que la evaluación tiene que ser objetiva y que cualquier grupo de evaluadores que haga el trabajo debe obtener los mismos resultados. La reproductibilidad de los resultados es uno de los criterios de validez en este modelo en el cual objetividad y fiabilidad son sinónimos. Al reducir a pocos indicadores los datos para hacer el cálculo coste-beneficio y descubrir la eficacia del programa, la complejidad del mismo queda muy reducida. Este modelo asume la perspectiva de los gobernantes e ignora los legítimos intereses de los participantes en los programas.

Sin embargo, a pesar de que tenían un gran interés para la administración, no bastó con las relaciones coste-beneficio, ya que no se consideraban una parte de los aspectos sustanciales de los programas, que presentaban un fuerte componente de estrategia política para este sector, con lo cual se hacía necesario que la evaluación considerara también las aportaciones del mismo programa a la población a que iba destinado.

D. Stufflebeam proponía de forma explícita la relación entre evaluación y la toma de decisiones, algo que hoy día es sobradamente aceptado. Este enfoque se ha denominado **toma de decisiones**.

D. Stufflebeam...

... definió la evaluación así: "... el proceso de diseñar, obtener y suministrar información útil para juzgar entre distintas decisiones posibles." (Stufflebeam, 1988.)

La evaluación se estructura en función de cómo se tomarán las decisiones, lo cual habitualmente hace referencia a las que tomarán quienes dirigen el programa y/o la Administración. El evaluador debe identificar el ámbito en que se toman y definir los criterios que se utilizan en las situaciones identificadas. Además, no planifica experimentos, sino que utiliza el método de muestreo y reúne la información, tanto en cuestionarios como en entrevistas. Se trata de un enfoque de tipo correlacional.

La audiencia principal...

... y casi exclusiva de la evaluación son quienes toman las decisiones, de manera que utiliza sus criterios e intereses como los auténticamente significativos, y pone en su boca preguntas como:

- ¿Qué partes del programa funcionan?
- ¿Cuál es la eficacia del programa en su conjunto?

Quizá lo más arriesgado de este enfoque es creer que aumenta considerablemente el uso de los resultados de la evaluación.

Patton, que ha contribuido significativamente a su desarrollo, ha enfatizado el factor personal para el uso de los descubrimientos de la evaluación. Por ello es crucial en este enfoque identificar las preguntas clave del programa, para centrarse en ellas. La evaluación informa de estas preguntas, que tienen distintas soluciones, pero quien toma las decisiones es quien realmente debe contestarlas.

M. Scriven desarrolló el enfoque conocido como **libre de objetivos**, que de alguna manera es una reacción al enfoque centrado en los objetivos.

Mientras que en el sector educativo...

... el modelo de Scriven es muy problemático por la dificultad de identificar el "producto" de la educación, en las organizaciones de consumidores, en cambio, ha sido muy bien recibido. El evaluador tiene que conocer la aceptación de los consumidores, desconociendo los objetivos de las empresas que elaboran el producto, lo cual permite generar criterios de evaluación de acuerdo con las necesidades de estos consumidores, que son su audiencia principal, y no los gestores de las empresas.

Para este autor, el evaluador no tiene que estar informado de los objetivos del programa, para evitar un cariz determinado en la búsqueda de resultados. El evaluador debe buscar todos los efectos que los programas generan, que pueden obviamente ser positivos y negativos, e intencionados o no. Este modelo, por lo tanto, se debe entender en el contexto que más preocupa al autor: evitar o reducir al menos los efectos de posibles carices personales en la tarea del evaluador, problema relacionado tanto con su independencia como con la objetividad del método que utiliza.

Para Scriven, la noción...

... tradicional de subjetividad y objetividad es cuantitativa y se fundamenta principalmente en el uso de mediciones que son reproducibles. La conceptualización que establece de este término en su modelo es ya cualitativa y se basa en la independencia de las diferentes influencias que interfieren en el juicio.

Al proponer que el evaluador desconozca los objetivos del programa para no verse condicionado por ellos, otros autores han interpretado que el evaluador los sustituye por los suyos propios. Sin embargo, Scriven resuelve esto centrando el resultado de la evaluación en el producto que el programa ha generado, que es más objetivable, y descargando nuevamente, de este modo, a la persona del evaluador.

El **evaluador** tiene que ser, por lo tanto, un experto capaz de evitar un cariz determinado debido tanto a sí mismo como a los múltiples factores externos, y debe perfilar su tarea como la de un detective que tiene que descubrir las claves de un hecho que ya ha pasado, es decir, determinar las conexiones causales en condiciones no experimentales.

A partir de este momento, el campo de la evaluación se diversifica mucho y surge una gran variedad de enfoques, como la **crítica de arte**, que es esencialmente de carácter empírico y cualitativa, y no necesariamente negativa, sino que ofrece el conocimiento de las cualidades y la calidad que el objeto posee para valorarlo.

Lectura complementaria

Para ampliar esta información, podéis consultar el libro:

M. Scriven (1991). *Evaluation: Thesaurus*. Beverly Hills: Sage.

Otro enfoque es el de la **acreditación profesional** o el **cuasilegal**, en que los aspectos considerados importantes por el público tienen que tratarse, de manera que aumenta considerablemente la participación respecto a otros modelos, y en general, si no todos, muchos puntos de vista quedan representados en el mismo.

Sin embargo, lo que ha provocado quizá más impacto y controversia ha sido el denominado *estudio de caso*.

El **estudio de caso** tiene como objetivo central mejorar la comprensión y el conocimiento del objeto evaluado a las audiencias de la evaluación. Quienes trabajan con este enfoque plantean la representación del programa como un proceso y ofrecen una visión del mismo tanto desde su perspectiva como desde la de todos los que participan en él.

Muchos autores han contribuido al desarrollo de este enfoque, incluyendo en la forma de hacer los estudios de caso su particular visión de la evaluación y el modelo evaluador.

Por ejemplo,...

... B. MacDonald considera la evaluación como una forma de servicio a la comunidad, a la cual facilita información sobre las características educativas del programa que se evalúa.

El valor básico es que los múltiples intereses que hay tienen que estar representados, y por lo tanto el papel del evaluador es en parte el de un intermediario de información entre los diferentes grupos de participantes en el programa. El objetivo de sus informes es servir al mayor número de audiencias posible, incluyendo a quienes toman decisiones.

Para R. Stake, los estudios de caso son narraciones descriptivas que muestran la complejidad de las situaciones del mundo real, y en que la comparación que se establece aparece de forma bastante implícita. Su modelo de evaluación (que denominó *respuesta*) es un intento de contestar a las diferentes audiencias lo que desean saber, para lo cual establece una interacción tanto con clientes como con participantes.

Desde el punto de vista metodológico, puede trabajarse con metodología cualitativa o cuantitativa, pero la primera es tan predominante que el enfoque queda claramente marcado por sus características.

En los estudios de caso...

... para la recopilación de información, se utilizan:

1. entrevistas;
2. observaciones;
3. estudios de documentos;
4. cuestionarios de preguntas abiertas.

La principal crítica a este enfoque es su subjetividad aparente, que comparte por otra parte con otros enfoques; sin embargo, quizá el principal problema que deben resolver quienes trabajan con el mismo es de qué manera se pueden representar los diferentes intereses que afectan al programa, y cómo equilibrarlos a la hora de ser representados.

2.2. Exclusión y complementariedad de métodos

Durante la última década, el método teórico de la evaluación se ha establecido en torno a la sustitución de paradigmas, mediante el uso de la terminología de los filósofos de la ciencia. Sin embargo, esta forma de pretender aclarar la elección metodológica y articular alternativas epistemológicas ha presentado algunas dificultades, como por ejemplo la falta de acuerdo sobre el significado del término *paradigma*, pero ha puesto de manifiesto el problema de la relación que hay entre paradigma y práctica.

Lecturas recomendadas

Para ampliar esta información, podéis consultar:

House, E.R. (1994). *Evaluación, ética y poder* (caps. 2 y 11). Madrid: Morata.
Stufflebeam, D.L.; Shinkfield, A.J. (1987). *Evaluación sistemática. Guía teórica y práctica.* Barcelona: Paidós-MEC.

Si entendemos por **paradigma** la visión del mundo o la organización de los principios que gobiernan esta percepción, el argumento de que la elección de paradigma debe hacerse teniendo en cuenta lo que mejor se ajuste al fenómeno estudiado parece, a simple vista, una forma pragmática y sensata. Sin embargo, también significa que podemos situarnos fuera de ellos y elegir a nuestra conveniencia, independientemente, teniendo en cuenta la naturaleza del problema, tal y como lo hacemos con el método que utilizamos. Esto sugiere que es posible definir los problemas independientemente de cualquier perspectiva epistemológica y que el eclecticismo es un objetivo al alcance.

De este modo, el propósito de algunos evaluadores y teóricos de integrar las metodologías cientificoexperimentales con las etnograficonaturalistas se convierte probablemente en un forcejeo en el momento del análisis y la interpretación de los datos, una vez que se ha aceptado la elaboración de un diseño común. Desde esta noción metafísica del paradigma, la elección estratégica del método y la práctica de la evaluación es muy difícil.

El cambio del paradigma científico al naturalista plantea un problema de relación entre la teoría y la práctica de la evaluación de programas, pero no un problema estructural. Esto se debe a que ambos tienen que plantearse las cuestiones políticas, morales, metodológicas y teóricas en relación con el programa que tiene que evaluarse.

De hecho, todos los evaluadores, en la práctica, se plantean cuestiones de utilidad, justicia, política, ética, relaciones entre las diferentes funciones y responsabilidad social, aunque puede argumentarse que las dos paradigmas definen estas cuestiones de manera diferente y que, por lo tanto, se llega a formas de actuación y compromisos también distintos.

Sin embargo, si se plantea el problema desde la perspectiva sociológica de los hábitos metodológicos, un paradigma tiene que describir la práctica de la evaluación, es decir, debe representar empíricamente las formas de hacer y de informar (las reglas metodológicas) de determinada comunidad evaluadora. De esta manera, el paradigma es la "tecnología" utilizada por la comunidad evaluadora para la resolución de los problemas que se plantean.

Analicemos el contexto histórico en que aparece la discusión entre los dos paradigmas, para que nos ayude a entender el fondo del problema. Las ciencias sociales se han encargado del estudio de la mejora de las condiciones de vida y de la comprensión de las tendencias, las circunstancias y las fuerzas que actúan en la sociedad. En su interior, encontramos una tradición en la cual el progreso social se ha estudiado teniendo en cuenta la práctica de experimentos. La experimentación proporciona la base para la planificación del cambio social, y son los planificadores en colaboración con los evaluadores quienes diseñan y tienen que analizar las innovaciones. Se aplica la lógica científica al estudio y el desarrollo de las innovaciones con el fin de producir un conocimiento preciso, válido y generalizable al servicio de la experimentación social.

En las "sociedades experimentales" se intentan innovaciones para solucionar los problemas, se establecen soluciones provisionales y se confía en la ciencia y la tecnología para hacerlo. Esta filosofía ignora la existencia de las fuerzas que han actuado históricamente, y es utópica porque se plantea "arreglar" la sociedad en su totalidad, lo que K. Popper ha denominado *ingeniería social*. El modelo de Rossi y Freeman, antes mencionado, aplica los métodos de la ciencia y contribuye a la planificación, el diseño y la implementación del programa, y en él, además, las mismas innovaciones se consideran también una ciencia aplicada.

Campbell criticó la...

... conceptualización de la evaluación como una ciencia aplicada, dando argumentos como por ejemplo que las medidas cuantitativas no pueden sustituir el conocimiento cualitativo, o cómo se pueden manipular fácilmente los objetivos de un programa convertidos en medidas, o la negación de que las verdades científicas son el producto colectivo de una comunidad en un tiempo determinado (Campbell, 1983).

El planteamiento experimentalista ha proporcionado una cultura del escepticismo basada en los valores del empirismo, ante el idealismo y el ejercicio administrativo del poder para manipular la información como más le convenga.

Cronbach criticó de forma pragmática y realista el planteamiento experimentalista. Aunque defendía que la evaluación es una ciencia aplicada, incorporó el componente político y pluralista a esta visión. La lógica del experimentalismo sugiere que es responsabilidad del evaluador decidir en qué condiciones, circunstancias y poblaciones el tratamiento practicado en un programa puede generalizarse, y éste depende de la representatividad

que tienen las muestras elegidas para hacer el experimento. Para Cronbach, la **generalización** es una cuestión de validez externa que se debe tanto a criterios metodológicos como políticos.

House y otros teóricos han criticado también la evaluación como aplicación de la ciencia, y enfocan su estudio en el comportamiento de la evaluación en las sociedades de mercado. Según él, la evaluación puede proporcionar recomendaciones creíbles, posibles y probables, pero no lo que es cierto (la verdad). La evaluación forma parte de los complejos procedimientos de decisión social y moral, y por ello la objetividad científica no trata con la profundidad que es necesaria el problema de las interferencias, con lo cual rechaza que la validez pueda tratarse estrictamente como una definición operativa.

En las evaluaciones practicadas con metodologías de tipo naturalista (en las cuales pueden incluirse los modelos iluminativo, de respuesta, democrático, etc.), las reformas e innovaciones tienen que hacerse sobre la base de la investigación. Se debe tener en cuenta que el objetivo de la misma debe ser informar de la toma de decisiones para conseguir mejoras sociales, y que se deben respetar los derechos y las perspectivas individuales de las personas, lo cual indica una orientación política de tipo liberal.

Las investigaciones naturalistas tienden también a evitar la construcción deliberada del control de las circunstancias sociales como medio de obtener nuevas ideas, y a cambio pretenden juzgar las acciones en su contexto. Por ello su objetivo es conseguir una mejor comprensión del programa y representar el trabajo que se está haciendo, informando de ello a las diferentes audiencias, para lo cual utiliza los métodos de otras ciencias sociales, como por ejemplo la antropología y la etnografía.

Enfoques como el de estudios de caso, sin embargo, tienen el problema de cómo hacer generalizaciones a partir de ellos, a pesar de que sus informes son muy accesibles. Parte del problema era quizá que la teoría de la generalización naturalista argumentada por R. Stake es, en opinión de algunos autores, el inicio de una teoría de la interpretación. Sin embargo, la evaluación debe proporcionar algunas generalizaciones sobre los programas y la experiencia que se ha producido. Estas generalizaciones se podrían resumir en:

1) Los juicios relativos a su naturaleza, procesos y efectos.

2) La transferibilidad de la experiencia y de las ideas desarrolladas.

Es cierto que la búsqueda de paradigmas alternativos se ha originado y se ha visto estimulada por una gran insatisfacción en el uso excesivo de los métodos

cuantitativos, y que históricamente la ciencia ha enfatizado enormemente la calificación, el tratamiento y la formalización matemática. Se ha entendido, incluso, que una ciencia llega a la madurez acompañada de un cierto grado de cuantificación, de tal manera que hay una convicción todavía muy extendida entre los estudiosos de que sólo los datos cuantitativos son válidos o de alta calidad.

Sin embargo, mientras que para las ciencias fisiconaturales la formalización matemática y el tratamiento estadístico han proporcionado resultados de gran utilidad para su avance, y cuando el objetivo de la ciencia es la predicción, al extenderse la investigación a las ciencias sociales, por el contrario, han surgido serias críticas. Entre ellas podemos mencionar la importancia de los contextos que afecta de lleno a la capacidad de generalización de los resultados conseguidos; la exclusión de significados e intenciones con la que trabajan los "científicos cuantitativos"; la imposibilidad de aplicar los datos generales a los casos particulares, o la consideración de que hechos y teorías no son apreciaciones de valor. La metodología cualitativa se considera actualmente adecuada para analizar hechos en los que el componente humano es la cuestión central.

Cronbach, al observar la forma de trabajo de la psicología, manifestó que había una diferencia importante entre los métodos: el experimental y el correlacional comparten lo que él denomina el trabajo de la ciencia, que consiste en hacer preguntas a la naturaleza.

El objetivo de los...

... correlacionistas es comprender y explotar la natural y presumible variación entre los individuos, mientras que los experimentalistas lo que intentan es trabajar creando condiciones en las cuales las variaciones se reducen.

Esto es para Schulman (1988) no sólo una forma de investigación que utiliza una estadística la cual recibe este mismo nombre, sino lo que hacen los investigadores que estudian la naturaleza cuando quieren entender las relaciones funcionales entre las variaciones de un conjunto de hechos y características y las variaciones de otros.

Los **experimentalistas** estudian las variaciones que ellos crean, y en su método los científicos cambian las condiciones para observar las consecuencias de los cambios que han introducido. Los experimentalistas dicen que sólo después de la introducción sistemática de cambios pueden distinguirse las relaciones causales entre hechos y características de las meras casualidades.

El problema reside quizá en el hecho de que se ha considerado la ciencia como una cuestión de método y metodología, en vez de una actividad intelectual. Históricamente, en cualquiera de los ámbitos de la ciencia, cuando determinados métodos se han revelado inadecuados para las nuevas formulaciones, han sido sustituidos por otros.

Para Ernst House,...

... los resultados obtenidos mediante ambos métodos son integrables en la naturaleza y el contenido del objeto que se evalúa.

Los métodos son importantes, pero lo que deben hacer es facilitar la tarea de evaluar y no distraer la atención de los puntos principales y sustantivos de la evaluación.

2.3. Fases del proceso de evaluación

Entendemos por **fases de una evaluación** las diferentes etapas por las cuales pasa el evaluador para llevarla a cabo, desde que se le encarga hasta que presenta su informe final. En general, puede decirse que las evaluaciones se encargan directamente a determinados evaluadores por su experiencia y competencia reconocida dentro del campo que hay que evaluar, en el uso de determinadas metodologías, etc. A veces se encarga a diferentes equipos de evaluadores hacer evaluaciones sobre un mismo programa para obtener estudios complementarios, los cuales, como vimos, pueden proporcionar diferentes enfoques.

Primera fase

Aunque ya hemos hablado de la necesidad de definir con detalle el objeto de la evaluación, es necesario remarcar que, en la primera fase del proceso, definir es una de las tareas esenciales, y que consiste básicamente en estos puntos:

a) definir las preguntas principales que la evaluación tiene que responder;

b) definir el objeto de la evaluación;

c) hacer el diseño de la evaluación;

d) acordar por escrito con la persona que financia y encarga la evaluación los tres puntos mencionados.

En esta fase la relación con el **cliente** es quizá un hecho determinante, aunque su importancia se mantiene a lo largo de todo el proceso.



En la primera fase es determinante la relación con el cliente.

En la interacción con el cliente, en esta fase, hay que responder a preguntas como las siguientes:

1) **¿Quién pide que se haga la evaluación?** Muchas veces la persona que se pone en contacto con el evaluador es el mismo director del programa objeto de la evaluación, pero son las personas de quienes depende administrativamente las que están pidiendo que se evalúe.

2) **¿Por qué y para qué se quiere evaluar el programa?** Ya hemos visto las diferentes necesidades por las cuales se evalúa un programa, pero es esencial para llevar a cabo su diseño saber cuáles son las consecuencias últimas que la evaluación puede tener. A veces, la evaluación precisa una pequeña investigación del evaluador, que puede encontrar la respuesta a lo largo del proceso.

3) ¿Qué tipo de información se necesita? En muchas ocasiones, el cliente tiene una idea muy precisa de lo que quiere para operar *a posteriori*. En ocasiones son datos que proporcionan una panorámica amplia de lo que ocurre en todos los lugares donde el programa se está implementando. En otras, se busca conocer los principales errores y/o éxitos del programa en su desarrollo para saber cómo modificar extensiones posteriores del mismo.



En lo que respecta al motivo por el cual se hacen las evaluaciones, consultad el subapartado 1.3. de este mismo módulo didáctico.

4) ¿Qué fuentes de datos se pondrán a su disposición? Es trabajo del financiador presentar la evaluación a los participantes en el programa para que accedan a darle la información que solicite, y también a poner a su disposición los archivos y documentos necesarios para investigar la historia y las directrices que marcan la política de desarrollo del programa concreto.

El evaluador debe tener...

... la oportunidad de entrar en el campo, poder definir el programa en acción y empezar a formular las primeras preguntas que surgen en la práctica del programa. Estas preguntas se utilizan como base para la elaboración del plan de acción que se presentará al cliente en la negociación definitiva de la evaluación.

5) ¿Cuándo hay que tener el informe de evaluación? Si la evaluación se solicita para tomar algún tipo de decisiones sobre el programa, el tiempo es importante. El evaluador debe trabajar en plazos reales.

6) ¿Con qué recursos cuenta la evaluación para llevarse a cabo? Conocer los recursos económicos con los que cuenta significa poder realizar la programación de las formas de reunir la información.

Lectura complementaria

Al introducirse en la parte de contenido más práctico, puede ser útil hacer referencia a un estudio de evaluación accesible, como el que se ha realizado sobre la enseñanza de las ciencias en el ámbito de enseñanza secundaria obligatoria de la reforma educativa, publicado por el **INCE** (1996) con el título *Currículo e innovación en la enseñanza de las ciencia*s, y que en el capítulo de metodología muestra el plan de acción y el diseño elaborado.

Segunda fase

Una vez que sabemos y tenemos definido el objeto de la evaluación, el siguiente paso consiste en la manera de estudiarlo. Estamos en disposición de hacer el diseño de la evaluación, que debe resolver las siguientes cuestiones:

1) ¿Para quién es la información? Esta pregunta nos lleva a pensar cuáles son las audiencias de la evaluación. Normalmente hay diferentes, y en cada evaluación en particular tiene que considerarse la importancia de cada una de ellas, pero de forma genérica pueden considerarse tres:

a) el cliente y financiador;

b) los que trabajan para el programa;

c) los que son el sujeto del programa;

d) los interesados en el programa.

Las audiencias de la evaluación se tratarán específicamente en un apartado posterior, pero son las que nos ayudan a centrar la información que necesitamos reunir. En general, puede ser útil realizar una lista detallada de las audiencias posibles y priorizarlas en función de la situación particular de la evaluación, es decir, seleccionar cuáles son las esenciales y cuáles las secundarias.

Lecturas complementarias

Para ampliar esta información, podéis consultar los libros:

Posavac, E.J.; Carey, R.G. (1992). *Program evaluation methods and case studies*. Prentice Hall.

Weiss, C.H. (1975). *La investigación evaluativa: métodos para determinar la eficiencia de los programas en acción*. México: Trillas.

2) ¿Qué información es la más conveniente? Una vez consideradas las audiencias es preciso preguntarse qué información le interesa a cada una, lo cual no es tan fácil de saber como parece a simple vista. Ha habido algunas evaluaciones en las cuales, para garantizar que un programa continuara siendo financiado, los evaluadores han resaltado sus puntos débiles, pues los políticos de los que dependía no podían permitirse un fracaso.

3) ¿Cuándo hay que tener el informe? Ya hemos mencionado la importancia de cumplir los plazos, pero lo importante es ser realista en cuanto a los tiempos en los que puede llevarse a cabo lo que se planifica. Depende de la forma de trabajo y la dedicación del equipo, pero cuando se está en condiciones de entrar en el campo (hecho el diseño y negociado los lugares de la recopilación de información), el tiempo total se divide en tres partes:



El tiempo es un factor importante para el evaluador.

a) la recopilación de información;

b) el procesamiento de la información;

c) la redacción del informe, en caso de trabajar sólo con un informe final.

La forma en que se organiza...

... la información condiciona el procesamiento que se haga de la misma; de hecho, los sistemas de organización de la información son en sí mismos una rama de la investigación. Para quienes trabajan con metodologías cualitativas, la organización de la información es una categorización de ésta, y habitualmente hay que cambiar la organización a medida que aumenta el número de datos.

4) ¿Cómo se reunirá la información? Las técnicas para la recopilación de la información se tienen que decidir en función de la información que se necesite, de las preguntas que se formulen en la evaluación y también de los recursos disponibles y de la metodología seleccionada.

5) ¿Quién reúne la información? Las personas encargadas de reunirla tienen que disfrutar de credibilidad entre los participantes en el programa, sobre todo cuando se trabaja con técnicas de tipo interactivo (observación, entrevista).

6) ¿Cómo se comunicará o publicará la información? Éste es un tema que gana cada vez más importancia. ¿Cómo conectar con las diferentes audiencias de la evaluación? ¿Cómo conseguir que el trabajo sea accesible a las personas que están interesadas en él? Éste es un problema tanto económico como de lenguaje, pero ya hemos remarcado la necesidad de utilizar el lenguaje de la acción y que el rigor de un informe no depende del uso de un lenguaje académico.



La segunda fase se centra fundamentalmente en la recopilación de la información. Tanto si se ha decidido utilizar un método de tipo experimental como otro tipo de metodología, es importante planificar con cuidado cómo almacenar la información.

El uso de ordenadores...

... facilita mucho la tarea de organización de la información. Hay un amplio número de programas al alcance de los diferentes métodos que aceleran bastante esta operación (R. Tesch, 1990).

Tercera fase

La tercera fase se centra en el procesamiento de la información. Nuevamente, se hará de forma distinta en función de la metodología utilizada, pero se trata

de la fase interpretativa por excelencia, en que los datos "brutos" pasan a ser pruebas del argumento que los autores de la evaluación desarrollan en el informe, en el cual se basan los juicios que se emiten del programa. Esta fase, como la anterior, no empieza de manera clara en este momento: el proceso empieza en las anteriores fases. Es una etapa en la que estudiar los datos, conocer sus significados e interpretarlos de forma correcta y adecuada es la clave del proceso, y en la cual hay que garantizar que aquellos relevantes para el argumento se han validado y reunido en diferentes fuentes de información y con distintas técnicas.

El objetivo fundamental...

... en la tercera fase es resolver organizativamente cómo encontrar una cita o un dato cuando se necesita. Si el sistema organizativo que se utiliza lo permite, puede considerarse que es aceptable.

2.4. Elaboración del informe de evaluación

La mayoría de las evaluaciones se hace sobre un solo programa, y muestra los efectos producidos por éste; si se utilizan grupos de control, quizá se podrá descubrir si el programa fue lo que causó estos efectos. Otras veces, lo que quiere conocerse es una comparación de los beneficios que distintos programas generan. Puede decirse que en la esencia de la evaluación reside el hecho de establecer una comparación entre diferentes programas, dentro de un mismo programa que se lleva a cabo en lugares distintos o sobre el servicio prestado a unos estudiantes antes y después de llevarse a cabo el programa. Sin embargo, la evaluación no da nunca todas las respuestas, pero sí que suele mostrar los puntos débiles del programa y las necesidades de cambio, lo cual es importante.

Lectura complementaria

Para ampliar la información sobre la construcción del argumento de la evaluación, podéis consultar:

E.R. House (1994). *Evaluación, ética y poder* (cap. 4). Madrid: Morata.

El **informe final** es para el evaluador el elemento en que se concreta la transacción que ha establecido con el cliente en primera instancia. El informe final es un documento que manifiesta todas las limitaciones que presenta una evaluación, algunas impuestas desde fuera como resultado de los procesos de negociación de los datos y del propio proceso investigador, y las impuestas por la concepción que el evaluador mismo tiene de la evaluación.



El informe de evaluación incluye las siguientes cuestiones: 

1) En primer lugar, un **canal de comunicación**. Algunos informes son un canal de comunicación con las audiencias internas del programa compuestas por individuos con los cuales se debe interaccionar en el proceso de investigación. Por ello, los informes no son sólo productos acabados y por lo tanto estáticos, sino dinámicos, porque se convierten en un elemento adicional de la recopilación de datos por los canales convencionales (entrevista, observación, etc.). Según cómo se dan a conocer, se produce en torno al informe una discusión en la cual la argumentación ante las diferentes interpretaciones de los hechos es, en sí misma, un dato importante para el evaluador.

2) Los informes son la representación de los hechos que se han producido, y de los pensamientos y las opiniones de los participantes. Esto significa, para el evaluador, un compromiso y al mismo tiempo un reto que tiene relación con el concepto de evaluación con el cual trabaja. El cliente, además de un juicio experto, busca también en muchos casos "escuchar a los participantes", a los cuales normalmente desconoce. Las consecuencias de que el informe sea una representación de la realidad a partir de los múltiples ángulos desde los cuales puede observarse el programa son lo que favorece, según nuestra opinión, que puedan introducirse modificaciones en el mismo, al disponerse de una visión plural y diferente a la que los distintos sectores interesados en el programa tienen.

Algunos autores...

... consideran la evaluación como un proceso de generación de recomendaciones, y éstas como el contenido propio de la evaluación (Patton, 1982).

3) Los informes suelen tener un apartado de recomendaciones, pero no es siempre así. En algunas evaluaciones no se piden, y en otras, sobre todo las que tienen estructura de estudio de caso, se espera que se generen en la interacción con el cliente. Sin embargo, esto depende en gran medida de cómo se haya formulado la evaluación. Los informes en que más énfasis se haga en las recomendaciones son los que están encaminados a la optimización del programa, pero el evaluador siempre debe tener presente que la evaluación tiene que cumplir un servicio tanto para clientes como para participantes.

La utilidad de la evaluación...

... se concreta muchas veces en el hecho de que a lo largo del proceso de la evaluación y del informe puedan obtenerse opciones de actuación distintas, que permitan intervenir en el programa y reconducirlo.

En esencia, la naturaleza del informe es la argumentación en que se basa la valoración que se recibe sobre el mérito y el valor (*merit and worth* en la formulación de M. Scriven) del programa. Es, por lo tanto, una representación holista del objeto que hay que evaluar, y que ofrece una comprensión más detallada y profunda a la vez que más global dentro de su contexto. La construcción del argumento afecta directamente a la validez del informe, y ésta no reside estrictamente, como ya veremos más adelante, en cuestiones como los instrumentos o métodos utilizados para la recopilación de la información, sino en que el argumento que se aduce como explicación se fundamente en un equilibrio de todos los intereses de participantes e implicados.

El formato del informe

La organización alrededor de los aspectos esenciales que configuran el programa suele ser la más habitual. Conviene conocer bien cuál es la estructura que facilita una mejor comprensión del programa al cliente, que es la audiencia que "está más interesada en el mismo", de tal manera que le sea útil a la hora de tomar decisiones.

La diferencia que hay entre los informes de tipo parcial, también denominados de progreso, y el informe final es importante: !

1) Los **informes parciales** no van necesariamente dirigidos a todas las audiencias, sino que es posible enfocarlos como un instrumento de trabajo, y utilizarlos como elemento negociador de datos y de recopilación de información adicional, por lo cual suelen ser de carácter restringido. En estos informes es especialmente importante que aparezca clara la diferencia entre los hechos o los descubrimientos, las interpretaciones, los juicios y las recomendaciones.

a) Los **hechos** se entienden como los datos empíricos y los descubrimientos básicos que se encuentran en la investigación.

b) Las **interpretaciones** son las explicaciones que se hacen sobre los hechos, y son también las especulaciones sobre las interrelaciones, las causas y los razonamientos de los hechos, es decir, el significado que se da a los datos.

c) Los **juicios** son las valoraciones que se hacen de los hechos, en el sentido de si son lo que se había proyectado, lo que se esperaba o lo que se considera como bueno o malo, deseable o no.

d) Las recomendaciones son lo que se sugiere para la actuación subsiguiente, es decir, lo que se estima que se tiene que corregir o mantener.

Lectura complementaria

Para ampliar esta información, podéis consultar:

M.J. Sáez Brezmes (ed.) (1995). *Conceptualizando la evaluación en España.* Alcalá de Henares: Universidad de Alcalá de Henares.

2) La construcción del **informe final** toca de lleno un problema que la evaluación comparte con la investigación educativa: la capacidad de generalizar los resultados que se obtienen de los datos reunidos. Campbell señala que, desde Hume, los filósofos y los investigadores han entendido que la generalización es un ejercicio de extrapolación que nunca puede justificarse de manera lógica. De hecho, ésta es una de las formas más comunes que se pide a los evaluadores que generalicen.

El informe final tiene que ser un documento dirigido tanto a las audiencias interesadas como a las audiencias naturales, para lo cual se hace necesario un lenguaje sin tecnicismos, pero preciso y riguroso, que permita formarse una idea muy precisa y detallada de lo que es el programa.

Ni los informes...

... sumamente detallados ni los puramente esquemáticos ayudan a comprender la problemática del programa; por ello, a veces se incluye un apartado con la historia que introduce las negociaciones que han permitido la financiación y la creación de este programa.

También es preciso, pues, que tengan un estilo literario que consiga interesar al lector, lo cual tiene que ver con el estilo narrativo y también con la extensión. El argumento se ha construido utilizando en el procesamiento todos los datos que se han recopilado, pero los que pueden incluirse en el informe son muy pocos. La selección resulta muy costosa, pero es preciso tener en cuenta que es recomendable incluir todos los datos en un anexo para que puedan ser consultados y para que las inferencias que se hayan producido puedan analizarse.

El informe final, aunque habitualmente se escribe para diferentes audiencias, tiene que ser único, y ésta es otra dificultad a la hora de escribir. Escribir distintos informes finales puede originar que se emitan juicios de valor sobre diferentes aspectos, y en consecuencia crear suspicacias legítimas sobre el carácter y la intención de los juicios en los diferentes informes. Esto afecta de lleno a la tarea principal de la evaluación, que es encontrar la forma adecuada de hacer público lo que sucede en los programas. Por lo tanto, es preciso hacer el propósito de que la participación en todos los ámbitos tiene que ser consecuente con las necesidades no tan sólo particulares de algún sector, sino de toda la comunidad de intereses.

El informe final suele tener carácter analítico, porque la extensión de los de tipo descriptivo es excesiva, lo cual significa que es necesario organizar la información en los puntos que se consideran más relevantes para la consistencia del programa y desarrollarlos como producto de la discusión que han originado.

El informe final tiene que incluir la metodología en la que se expone cómo se ha trabajado, las particularidades de las técnicas utilizadas y las negociaciones que la recopilación de datos ha generado, y el grado de participación alcanzado. Se debe explicar la elección de la metodología utilizada y las atribuciones con que el evaluador cuenta cuando empieza el estudio, para que el lector tenga una idea más precisa de las ideas previas del autor y, de este modo, se le reconozca el cariz personal, y para darle al informe un mayor grado de autonomía.

En resumidas cuentas, un informe final tiene que describir el programa y su contexto, justificar los criterios que se han seleccionado para establecer los juicios emitidos, mostrar los datos base del argumento, describir la metodología con que la evaluación se ha llevado a cabo y prestarse a obtener conclusiones o recomendaciones que permitan la toma de decisiones en todos los niveles del programa.

3. Evaluación, seguimiento y optimización de programas educativos

3.1. Audiencias y pluralidad de intereses

La **audiencia de una evaluación** son los grupos de individuos que pueden utilizar sus descubrimientos o resultados y reaccionar en contra de los mismos. Toda evaluación tiene diferentes audiencias y, por lo tanto, su presentación pública se debe planificar cuidadosamente para atenderlas a todas.

Consumidores y audiencias

En una sociedad de mercado como la actual, el término **consumidor** no es ajeno al concepto de **audiencia**, porque los consumidores de un programa, un servicio o un producto son las personas que de forma directa o indirecta lo utilizan, de modo que los consumidores de una evaluación pueden considerarse sus audiencias.

¿Cuáles son habitualmente las audiencias de una evaluación? O mejor dicho, ¿a quién le interesan y le son útiles los resultados y los descubrimientos de una evaluación? Pueden agruparse de la siguiente manera:

a) los que financian la evaluación;

b) los organismos o las instituciones de los cuales depende el programa que se evalúa, que no se identifican siempre con los anteriores;

c) los directores concretos responsables del programa;

d) el personal que trabaja para el programa;

e) la población a la cual van dirigidas las actividades del programa;

f) los que se ven afectados por el programa;

g) los profesionales que, aunque no participen en el programa, tienen relación con él o interés académico o profesional en el mismo.

Según el objetivo de la evaluación, unas evaluaciones pueden considerarse más relevantes que otras; así, el cliente que la encarga para tomar decisiones se considera una audiencia clave a la cual, sin duda, el informe le tiene que resultar útil, pero también lo es a los planificadores y directores locales, ya que su actuación está claramente sometida a escrutinio, y sus descubrimientos y resultados pueden tener valor para enderezar algunas actividades. A todas estas audiencias, y a los que trabajan para el programa, se les denomina **audiencias naturales**, ya que todos, de una manera u otra, participan del programa.

Se denomina ***stakeholders*** o **audiencias interesadas** a todos los demás, es decir, aquellos que están interesados en el programa de forma directa o indirecta,

ya sea porque su trabajo profesional tiene relación con él o porque son profesionales de otras comunidades que tarde o temprano se verán afectados por el mismo, o porque actúa indirecta e involuntariamente sobre algunos colectivos.

Las audiencias cambian durante el proceso evaluador

Sin embargo, la prioridad de las diferentes audiencias a lo largo del proceso evaluador puede cambiar. Es decir, probablemente al inicio de la recopilación de información las autoridades de ámbito local son las que están más interesadas en los descubrimientos de la evaluación, pero más adelante lo están también los directores concretos de los proyectos, y finalmente los más interesados son quienes deben tomar las decisiones con respecto a las líneas maestras que orientarán las directrices de la política del programa. También puede ocurrir que a lo largo de la evaluación surjan algunos grupos interesados con los cuales no se había contado inicialmente.

Si observamos la evolución del concepto de evaluación, vemos que el papel de las audiencias ha llegado a ser más importante en los diferentes modelos, hasta convertirse en un elemento definitorio para la elaboración del diseño.

Consultad el subapartado 2.1. de este mismo módulo.

Por ello, la pregunta más importante que hay que formularse en relación con las audiencias es la siguiente: ¿qué necesidades tienen y qué tipo de información se debe dar a cada una?

Según Raizen y Rossi (1981), dos hechos resultan esenciales en respuesta a esta pregunta:

1) Hay diferencias importantes entre los usuarios potenciales y los financiadores de la evaluación, y éstos no siempre deben ser considerados la primera audiencia.

2) En la formulación de las preguntas clave que guían la investigación de la evaluación también se deben tener en cuenta otras audiencias.

Parte del problema del uso de los informes de evaluación se debe al hecho de que la difusión del conocimiento generado no llega a todas las personas que tienen nivel ejecutivo. También hay una opinión generalizada de que el uso de los resultados tiene que hacerse a corto plazo. Sin embargo, en opinión de los autores mencionados, la difusión de descubrimientos y resultados se debe prolongar por un tiempo, procurando que los informes puedan conseguirse fácilmente, porque a veces es bastante difícil.

El lenguaje y la extensión de los informes son también importantes para que éstos sean accesibles. Para conseguirlo, pueden hacerse algunas cosas:

– Identificar, desde el comienzo del estudio, cuál es la primera audiencia.

– Preparar contactos entre los evaluadores y los que pueden utilizar sus resultados y niveles ejecutivos, lo cual, además, es posible que garantice una cierta fidelidad a los resultados de la evaluación.

– Escribir artículos que ayuden a aproximar las preguntas particulares que diferentes grupos se hacen sobre el programa.



En lo que respecta a la publicidad de los informes, surgen fácilmente dos preguntas clave. ¿Cuáles son los derechos que tienen quienes financian una evaluación sobre el informe final? ¿Y cuáles son las negociaciones de los evaluadores para difundir los resultados? Si se quiere optimizar el funcionamiento del programa, los resultados deben ser conocidos por los que trabajan para el programa y por los que se ven afectados por su uso. Cuando nos referimos a un programa educativo, los estudiantes son habitualmente los más afectados, aunque muchas veces, por su edad, no se consideran una audiencia adecuada. Sin embargo, sí que pueden serlo sus representantes directos: los padres.

Los resultados...

... de la evaluación tienen un marco adecuado para su discusión en las revistas profesionales, en las cuales se espera que se planteen controversias, y aunque éstas se consideren audiencias de tipo secundario, son lugares donde es posible comunicarse de forma diferente de como se haga con los informes.

Una norma razonable puede ser la que proponen Rossi y Raizen, según los cuales el informe se debe hacer llegar a todos aquellos que en función de representar a los beneficiarios del programa reclamen conocer sus resultados. Si la evaluación está financiada con fondos públicos, es el público el que tiene derecho a conocer sus descubrimientos y resultados. De este modo, la evaluación está tratando con el derecho a conocer que tienen los ciudadanos en las sociedades democráticas.

¿Resulta por ello pertinente, ante esta situación, preguntarse de dónde provienen los criterios para evaluar un programa? Ha sido el reconocimiento de la evaluación como un acto público lo que ha hecho que el establecimiento y la argumentación de los criterios de valor lleguen a ser un tema central de estudio en la actualidad.

Volviendo a los modelos de evaluación estudiados, los objetivos del programa son los que se utilizan como la fuente de criterios de valoración (modelo por objetivos) o bien la utilidad para la toma de decisiones (modelo de Stufflebeam), pero M. Scriven, cuando argumenta su modelo, plantea que son los propios objetivos del programa los que tienen que evaluarse, y que el interés del director del programa puede no ser relevante si lo que se evalúa es un producto.

En los programas educativos...

... hay múltiples criterios de valor a menudo conflictivos; por ejemplo, la proporción entre historia y matemáticas en un currículum. La gente no quiere lo mismo de los programas públicos, y difiere tanto en sus valores como en sus intereses; es decir, un programa puede ser bueno para unos, y no serlo para otros.

Algunos evaluadores eligen la forma tradicional de medir la puesta en marcha de un programa y sus resultados, porque asumen que los objetivos reflejan los intereses de participantes y *stakeholders* y la existencia de consenso sobre los valores del programa. La ausencia de este consenso ha hecho que otros evaluadores hayan tratado el problema dando protagonismo a clientes y participantes, tanto a la hora de definir lo que es importante para la evaluación y pedirles sus opiniones como en el momento de atenderlos como audiencias e, incluso, haciéndolos partícipes de algunas decisiones del estudio. Sin embargo, constatar la existencia de múltiples intereses no soluciona el hecho de cómo conjugar esta multiplicidad para emitir juicios. Los intereses a los cuales sirve la

evaluación y la forma en que éstos se representan son asuntos cruciales. Por ello, las "audiencias interesadas" (*stakeholders*) se convirtieron en un concepto común, y representar sus puntos de vista fue una práctica aceptada.

Según Weiss,...

... el enfoque basado en los *stakeholders* "promete un mayor grado de justicia en la evaluación a través de incrementar el rango de información recogida, al dar más voz a los grupos locales y al democratizar el acceso a la información, pero sigue sin resolverse cómo armonizar los conflictos de intereses entre los distintos grupos." (Weiss, 1975.)

Este modelo se basa en una teoría de la democracia en la que ésta es un equilibrio pluralista elitista, que descarta los juicios normativos y defiende que en el sistema actual la competición entre las partes y la presión de los grupos condiciona y perfila la función democrática de equilibrar las demandas políticas que son variadas y cambiantes. Evaluaciones (*Push/Excell*) realizadas con este enfoque muestran limitaciones, como por ejemplo que los intereses de los *stakeholders* pueden trabajar en contra de los participantes porque la influencia de los evaluadores no es la misma que la de éstos, que tampoco son un grupo homogéneo con la misma capacidad de influencia.

Lectura complementaria

Para ampliar esta información, podéis consultar el libro:

E.R. House (1993). *Profesional Evaluation. Social Impact and Political Consequences* (cap. 7, págs. 120-123). Newbury Park, Londres y Nueva Delhi: Sage.

3.2. La negociación en la evaluación

La negociación en la evaluación es un concepto vinculado al reconocimiento de la pluralidad de intereses y a la idea de que el papel de la evaluación es contribuir a una progresiva y suave adaptación de las actividades sociales y a estructurar el cambio de las condiciones y los ideales.



Consultad el subapartado 2.2. de este mismo módulo.

Tiene su origen en la crítica que Cronbach hizo al experimentalismo, que lo llevó a formular la naturaleza política de la evaluación, y no como una cuestión estrictamente metodológica.

La opinión de Cronbach

Cronbach criticó la ortodoxia del experimentalismo, que se entendía como la forma de trabajar con hipótesis al estilo de las ciencias experimentales y con la idea de que los objetivos representaban una situación de consenso. Propuso, por lo tanto, una nueva conceptualización de la función social de la evaluación y de los problemas metodológicos, a pesar de que continuaba defendiendo los valores de la utilidad del modelo de la ciencia aplicada, porque según él el desinterés de la investigación científica ayuda a establecer la credibilidad y presiona para actuar razonablemente.

Para Cronbach, lo que se estaba pidiendo a la evaluación era una información que sirviera de base para una negociación política, en lugar de una información destinada a indicar una decisión correcta. Consiguientemente, plantea el papel del evaluador como el de un "científico del público" (*public scientist*) que sirve sus intereses hasta donde su trabajo se lo permite, con la responsabilidad de representar todos los sectores de la comunidad de forma no partidista y desinteresadamente. Entendía que las políticas sociales en Estados Unidos se estaban configurando por medio de una comunidad de intereses plural, caracterizada por la negociación y el compromiso con las decisiones resultantes, en la cual los políticos articulaban los valores e intereses del público.

En su teoría se percibe una reacción ante el dogmatismo metodológico y, por consiguiente, propone una nueva formulación del concepto de validez y utilidad. Para los experimentalistas, son los diseños experimentales mismos los que garan-

tizan la validez (interna) de la evaluación y, sobre todo, de las conclusiones que se obtienen de ella. Cronbach redefinió los procesos de elaboración de inferencias desde los datos obtenidos, del mismo modo que la validez interna (reproductibilidad de las inferencias) como la que lleva a las conclusiones de la investigación, y la validez externa (extrapolación de los datos) como la que lleva a las conclusiones que se hacen en el terreno de la aplicación.

Campbell...

... contribuyó a la polémica señalando que la validez interna y la validez externa, según la definición dada por Cronbach, dependen una de otra.

House y Cronbach están de acuerdo en que hasta cierto punto una teoría de la evaluación trata de la interacción política y de la determinación de los hechos de los cuales se deben pasar cuentas. Por ello, House interpreta que la validez de la evaluación no puede separarse de las circunstancias y consideraciones políticas y sociales. House hace el análisis de la validez centrándose en el componente de intencionalidad social de la evaluación. Argumenta que desde este punto de vista se discriminan ciertas alternativas y ciertos intereses, y se favorecen los más fuertes, que están representados por las posiciones gubernamentales. Éstas operan con una definición individualista de los intereses y una atribución pluralista de la distribución del poder que ignora la diferencia que hay en las costumbres, las tradiciones y organizaciones administrativas, sociales y políticas.

House critica la ética utilitarista que impregna la evaluación, lo cual la hace más presente en las actuales desigualdades, y de este modo sirve al *status quo*. Desarrolla, por lo tanto, una teoría de la justicia de carácter pluralista e intuitiva que denomina *naturalista*, que tanto evaluadores como implicados tienen que observar, y que resume en cuatro valores principales:

– igualdad moral,

– autonomía,

– imparcialidad,

– reciprocidad.

Los cuatro sostienen el "acuerdo justo", cuya consecución tiene que dirigir la evaluación. Ninguno tiene la primacía sobre los demás y se deben priorizar según las circunstancias. De hecho, algunos recuerdan a los que Campbell propone para la sociedad experimental: relaciones sociales marcadas por la honestidad, la sinceridad y la no aceptación de dogmas; rendición de cuentas mutua; buena voluntad hacia el cambio; y respeto por los procesos.

De las formulaciones propuestas por estos dos autores se deduce que la metodología de la evaluación presume ciertas formas de relaciones sociales y anticipa ciertos tipos de vida política. La metodología de la evaluación sugiere una nueva teoría social.

B. MacDonald señala que las opciones del evaluador no tienen que limitarse a elegir entre reforzar o alterar las relaciones de poder existentes, aunque deja sin resolver la articulación de las condiciones políticas y éticas para practicar la evaluación de forma independiente al perseguir los valores democráticos. Para algunos teóricos, este enfoque es una estrategia que tiende a igualar las relaciones de poder en que están inmersos los programas, al tratar de hacer accesibles los informes y los métodos a las audiencias no especialistas, en un intento de democratizar el acceso al proceso y al producto de la evaluación.

El modelo defendido por B. MacDonald...

... es quizá el más político porque se centró en la exploración de los papeles políticos no partidistas, pero radicales, de la evaluación.

Para algunos modelos de evaluación la negociación se convierte, de forma explícita, en un elemento fundamental de su conceptualización, con valor operativo en su metodología, y se define como "una estrategia de intervención de carácter polivalente y por tanto adaptativa a contextos diferenciados según sea su objeto relevante" (A.J. Carretero, en M.J. Sáez, 1995).

En estos modelos, la evaluación tiene como intención colaborar a hacer más eficaz el funcionamiento de la organización social, por medio del establecimiento de modelos participativos basados en la comprensión de las situaciones y de la responsabilidad propia, y promover con ello sistemas que consigan aprender mediante formas organizativas colaborativas y canales de comunicación fluidos y no jerarquizados. El papel del evaluador se perfila, por lo tanto, como el de un intermediario que facilita la información que sirve para poner en comunicación las partes en la negociación, y como base para alcanzar acuerdos que se fundamenten en ella, ya que tiene que haber sido independientemente recabada por los evaluadores. La utilidad de la evaluación estará, pues, en función de su capacidad para generar las condiciones tanto de contenido como de procedimiento que permitan un acuerdo sobre si el programa tiene que modificarse, difundirse o suspenderse.

En este modelo se señalan tres ámbitos para la negociación: 

1) **Ámbito metodológico:** de procedimientos, que permite que los participantes en el programa evaluado intervengan tanto en las descripciones, las interpretaciones y la organización del propio informe evaluador como en la reconstrucción de los interrogantes, a causa de que los datos de la evaluación están inmersos en los significados y las formas de conocimiento de los participantes. Se debe observar como estrategia de descubrimiento que se desarrolla en el contexto epistemológico de la evaluación, en la generación de conocimientos, el contraste de interpretaciones y la validez de los datos.

2) **Ámbito ético**: en lo que concierne al uso de información (confidencialidad/publicidad), que busca los acuerdos normativos con respecto a los derechos de los evaluados y principios de transparencia y representatividad en el uso de la información. Estas competencias evaluativas de ámbito normativo y deontológico, es decir, los derechos y los deberes de evaluadores y evaluados, pueden concretarse en los tres siguientes apartados:

a) El derecho a la propiedad de la información, que corresponde a quienes la proporcionan en su colaboración con la evaluación, su crítica, etc.

b) El derecho de los individuos como audiencias interesadas y naturales a tener conocimiento de los programas que actúan y condicionan el espacio político y público y los descubrimientos y resultados de las evaluaciones que se hacen.

c) El derecho de todos los grupos de interés a estar representados en la evaluación y el deber de los evaluadores de garantizar su independencia e imparcialidad al tenerlos en cuenta.

3) Ámbito político: de la representación de intereses y valores, como estrategia de legitimación que configura el espacio político de la evaluación en tanto que se ocupa del sistema de valores que orientan los juicios de la evaluación y la distribución y correlación de fuerzas dentro de los diferentes intereses en el programa y en la evaluación.

3.3. Evaluación interna y externa

El incremento del uso de evaluaciones internas frente a las externas significa un cambio muy importante en la práctica de la evaluación, y se produce cuando ésta empieza a ser reconocida como una herramienta para la gestión de organizaciones.

Las evaluaciones internas...

... en Estados Unidos han crecido muy rápidamente durante los últimos años, a causa tanto de recortes en los presupuestos como de un cierto desencanto con los resultados de las evaluaciones externas.

De hecho, la evaluación interna comporta una forma de flexibilizar el funcionamiento de todo tipo de organizaciones, y por lo tanto también una vía para optimizarlo.

Algunos autores...

... dedicados al estudio de organizaciones empresariales definen la evaluación así: "La evaluación es sobre todo un proceso inmerso en la gestión de la organización, la toma de decisión y la planificación [...] La evaluación de programas puede entenderse como un aspecto integral del diseño y el desarrollo organizativo."

La evaluación interna es la que realizan los miembros de una organización que no tienen responsabilidades en el diseño y desarrollo del programa que quiere evaluarse, a diferencia de los evaluadores externos, que sí las tienen. Los evaluadores internos tienden a estudiar los problemas que interesan a quienes gestionan el programa, mientras que los externos habitualmente se encargan de estudiar problemas relacionados con los programas, que interesan a los políticos, los financiadores, el público en general, etc.

Para algunos autores,...

... la evaluación interna es una forma de investigación-acción (Adelman, 1981) que da apoyo al desarrollo organizativo y el cambio planificado, de manera que los evaluadores internos son a menudo responsables no sólo de analizar los problemas y hacer recomendaciones, sino también de corregir dificultades y aplicar soluciones.

La forma en que la evaluación interna trata los objetivos, los principios, las audiencias y las relaciones con los miembros del programa evaluado es diferente de como lo hace la evaluación externa.

Se ha debatido durante mucho tiempo qué tipo de evaluaciones es el mejor para un programa, si las internas o las externas, pero ninguna tiene el monopolio de las ventajas. Por el contrario, hay una serie de factores que deben con-

siderarse y en función de ellos decidir cuál es la más conveniente para la situación concreta. Estos factores son:

1) **Confianza de quienes la solicitan**: se basa en parte en el conocimiento de las destrezas básicas de la evaluación, y también en la capacidad de los evaluadores de acercarse a la realidad del programa, algo que en principio supone una mayor dificultad para los externos.

2) **Objetividad y distanciamiento**: en general, esta característica la cumplen mejor los externos porque están fuera de la problemática y no dependen de las relaciones internas, aunque los externos, en ocasiones, para conseguir nuevos contratos pueden suavizar los resultados.

3) **Comprensión del programa:** se trata de un factor fundamental, como ya vimos anteriormente, para una buena evaluación. En este sentido, los evaluadores internos están mejor situados que los externos, pero éstos también pueden tener un excelente conocimiento de la problemática que se crea en el programa, si se les da acceso a las fuentes de información.

4) **Potencial de utilización**: la utilización de los resultados requiere un papel activo de los evaluadores para convertir los datos de la investigación en recomendaciones de política general, y tal vez los evaluadores internos tengan más posibilidades de formularlas para el programa y defenderlas. Sin embargo, los externos disfrutan de más autoridad y credibilidad y quizá se les preste más atención.

5) **Autonomía**: los evaluadores internos suelen dar por conocidos los supuestos fundamentales y los esquemas organizativos de los programas que evalúan, y tienden a hacer la evaluación sin cuestionar este marco. Los externos tienen más facilidad para situarse en una posición que les permita tener una visión más amplia, y aun respetando la formulación de la problemática planteada por el programa, es posible que de sus resultados puedan obtenerse conclusiones que vayan menos encaminadas a buscar soluciones provisionales y más a reestructurar el programa a fondo.

La evaluación interna empieza cuando la institución u organización que promueve el programa es consciente de las necesidades informativas que tiene y de sus beneficios estratégicos. Permite al programa tutorizar sus operaciones y ayudar a los gestores a entender las actividades de la organización, y hace que éstas sean más manejables.

Cuando el objetivo es rendir...

... cuentas sobre el programa a los legisladores, accionistas o financiadores, los evaluadores externos tienen un mayor grado de credibilidad, porque aunque los internos puedan actuar de forma independiente y autónoma, no dejan de ser empleados de la institución, condicionados por el lugar que ocupan en la estructura.

Entre sus actividades está la priorización y valoración de necesidades, consistente en revelar los desfases entre los resultados obtenidos y los deseados; realizar un seguimiento del programa ofreciendo retroalimentación *(feedback)* sobre su operatividad y los posibles problemas; y hacer la lectura de su impacto (A. Love, 1991).

Los evaluadores internos están sujetos a todas las presiones que hay en la organización y, con frecuencia, son sólo una herramienta de los administradores, con los cuales suelen identificarse en lo que respecta a sus intereses. Esto no quiere decir que éstos no tengan necesidades legítimas de información para gestionar adecuadamente la organización que los evaluadores, sin duda, deben aportar. Otro problema, menos obvio quizá, sería que las evaluaciones internas generan rutinas, y convierten las tareas de evaluación en repetitivas, lo cual en ocasiones hace que se pierda su principal significado.

3.4. Autoevaluación

Para R. Tyler, la intención de la evaluación educativa era mejorar la enseñanza haciendo que el productor y el consumidor de la información y del conocimiento fueran la misma persona. L. Stenhouse, tomando alguna de las ideas de Tyler, planteó que la investigación llevada a cabo por el profesorado tenía un papel fundamental en la mejora de la calidad de la enseñanza. Para él, la clave consistía en entender la enseñanza como una forma de investigación, lo cual ofrecía además una solución al problema de la relación entre teoría y práctica para la investigación educativa. Para Stenhouse, el currículum era una recomendación política expresada en el marco de la acción. Por este motivo, era necesario integrar currículum y evaluación en un modelo, y perfilar así el papel investigador del profesorado y no tan sólo como creador del currículum. Su teoría daba prioridad al juicio profesional y a los profesores.

Según Norris, en Inglaterra, al final de los setenta, los que se dedicaban al desarrollo del currículum y la evaluación sugirieron que los programas definidos de forma centralizada necesitaban iniciativas locales, como una forma de darles cierto apoyo. La importancia del apoyo local encontró formas de expresión como el profesor-investigador, la red investigación-acción de aula, y la formación permanente basada en los centros escolares y en el desarrollo curricular. Estas formas de desarrollo pueden entenderse como estrategias organizativas y políticas de investigación curricular y de reformas arraigadas en las bases.

La autoevaluación...

... de centros escolares e instituciones, y el crecimiento de las actividades de evaluación formativa son también formas de expresión de los valores locales.

Desde hace muchos años, las asociaciones de profesionales (médicos, abogados y educadores) han llevado a cabo evaluaciones sobre cómo se forma a los profesionales del sector. Este enfoque es un modo de autoevaluación realizada por profesionales del sector y, en muchas ocasiones, por los propios miembros de la organización objeto de la evaluación. Este modelo ha inspirado la mayoría de las evaluaciones que se llevan a cabo en las universidades, y asume que no hay nadie mejor que los profesionales del área para juzgar el mérito y el valor de la institución, de tal manera que la formulación de criterios la hacen ellos mismos. Este modelo tiene poca credibilidad fuera de las asociaciones y,

en especial, hoy día, esto se debe a la presión de las organizaciones de consumidores.

En este enfoque se plantean dos problemas importantes:

1) Si los informes son confidenciales.

2) Si los informes son justos y tratan de la misma manera a todos los grupos y estamentos implicados en la evaluación.

En Inglaterra, a mediados de los años ochenta, se aplicaron mucho los modelos mixtos de evaluaciones formativas llevadas a cabo en el ámbito local y sumativas en el ámbito nacional. Los que trabajaban desde la perspectiva local lo hacían a partir del planteamiento del desarrollo profesional del profesorado. P. Holly ha denominado a este modelo de retroalimentación formativa **evaluación profesional**. !

La autoevaluación...

... con intenciones formativas se enmarca en ciertas concepciones de la mejora de la enseñanza y apela a un grupo de profesionales que se siente muy vulnerable a la crítica.

El hecho de que la autoevaluación con intenciones formativas pueda o no proporcionar una alternativa creíble a la evaluación externa es más que cuestionable, pues significa la evaluación de servicios públicos en un contexto altamente politizado y con muchos clientes e intereses. La evaluación profesional como modelo parece pensada para un sistema social que disfruta de acuerdos en sus objetivos, valores y criterios de éxito y con unos profesionales que pretenden alcanzar lo mismo. Sin embargo, este mundo “ideal” difícilmente puede ser la base de la estrategia política para la evaluación y los evaluadores en el mundo actual.

Muchas de las discusiones sobre la evaluación de programas y políticas llevan a considerar las clases y lo que informa la práctica de los profesores, con el fin de conocer el modo en que la evaluación puede ayudarlos a cambiar. La importancia del papel del profesor en las teorías de desarrollo curricular y la evaluación es una característica que se debe en parte al énfasis puesto en la autoevaluación como respuesta a la presión de rendición de cuentas de los años setenta.

J. Elliott recomendaba...

... el desarrollo de un sistema profesional y democrático de rendición de cuentas, aunque para él la autoevaluación no era sólo una forma de rendir cuentas, sino un instrumento para que los profesores mejoraran su forma de dar clase y de conocimiento, con nuevos modos de profesionalidad.

Mientras que después de Tyler la tradición americana se impregnó de racionalismo científico, según el cual el profesor es un trabajador en un sistema productivo sujeto a la evaluación de la puesta en funcionamiento y el control de calidad, la idea de Tyler de mejorar la calidad de la enseñanza a partir de la estimulación de los profesores para que evaluaran su propia práctica y sus iniciativas curriculares se perdió. La tradición inglesa cambió a lo largo de los años ochenta, cuando se centralizaron las innovaciones y la evaluación reflejó la relación entre productores y consumidores de la información de la división del trabajo.

Si la mejora de la calidad de la enseñanza se centra en el aula, las iniciativas de autoevaluación pueden tener cabida. Sin embargo, si se enfoca en el marco del sistema educativo como mejora social, la autoevaluación, hasta ahora, no ha tenido mucho espacio, aunque su potencialidad en los procesos de cambio educativo continúe siendo defendida.

Elliott ha señalado...

... que el énfasis puesto en la rendición de cuentas de las escuelas sólo ha servido para que el poder que éstas tenían fuese a parar a otras instituciones. Simons sostiene, en cambio, que ha dado más poder a la gestión escolar y que la idea de Stenhouse es ahora más difícil de implementar que cuando la propuso en 1965.

4. El evaluador como profesional

4.1. La profesión del evaluador

Para entender la forma en que la evaluación se ha convertido en una profesión, es preciso empezar por preguntarse por qué razones la evaluación existe, analizar la evolución histórica de los modelos que ha ido adoptando y sus aspectos institucionales.

La evaluación ha aparecido en países de capitalismo avanzado como Estados Unidos, Canadá y los países del norte de Europa, además de otros, en un momento en el cual los gobiernos se distanciaban de sus apoyos tradicionales y de la población. Para gobernar, han necesitado apelar a ciertas formas de autoridad de diferentes maneras, pero tienden a hacerlo con la aceptación de la población. De este modo, la legitimación de las políticas no es ya una forma de control o imposición de ideas a la población, sino una justificación, mediante razonamientos y evidencias, de que los intereses de la gente pueden estructurarse en este contexto social.

Así pues, la evaluación formal se ha ido convirtiendo en una forma de autoridad cultural, en las sociedades en que el sistema de valores es cada vez más complejo y variado.

Según Ernst House,...

... la autoridad cultural de la evaluación se basa en la persuasión, que se manifiesta en la probabilidad de las descripciones de la realidad y en la validez de los juicios de valor. En esta dinámica social, la evaluación formal empieza su proceso de institucionalización (E.R. House, 1993).

Sin embargo, ¿qué es lo que la convierte en una profesión? La legitimación de la autoridad profesional implica tres aspectos:

- Que el conocimiento y la competencia profesional se hagan válidos para la comunidad (en este caso, los padres).
- Que este conocimiento tenga una fundamentación científica y académica.
- Que el juicio y el consejo profesional se orienten hacia valores sociales importantes.

Cualquier profesión...

... tiene el monopolio sobre un conocimiento que es importante para el funcionamiento de la sociedad actual y está compuesto por principios abstractos y experiencia práctica.

Cuando el conocimiento se estandariza, se empieza a producir una separación entre su parte teórica y la práctica. Antes de esto, los individuos se formaban en el conocimiento práctico con un evaluador de experiencia, y se relegaba a un segundo plano el conocimiento teórico. El desarrollo del conocimiento es lo que hace que se separen ambos campos, que el teórico crezca y se especialice en los temas sustanciales y que la universidad se dedique a la investigación y las carreras académicas.

Los **evaluadores** forman actualmente una comunidad relativamente grande con una pequeña parcela de poder, con la intención de contribuir de manera sustancial a que se lleven a cabo programas de calidad, de tal manera que quienes quieren dirigir un buen programa requieran el servicio de la evaluación. La consolidación de esta función comporta una lucha que sólo puede ganarse si todos los que trabajan como evaluadores son capaces de dejar de lado sus ideas particulares y centrarse en los aspectos prioritarios:



Según Perkin,...

.. todas las profesiones aparecidas en los últimos cien años han tenido una base en la educación formal que ha contribuido a la formación de la subcultura profesional, y las universidades han sido las instituciones encargadas de formar a los profesionales y producir el conocimiento profesional (Perkin, 1989).

Facultad de Biología de la Universidad de Barcelona.

1) La necesidad de una maquinaria y una moral profesional que pueda influir sobre los evaluadores y en la posteridad.

2) La profesión tiene que establecer acuerdos internos sobre lo que son las prácticas recomendables, generar conocimiento público y presionar a sus clientes para que busquen servicios que les sean beneficiosos.

3) También es preciso que quienes la ejercen sean competentes.

Los evaluadores se ven a sí mismos neutrales respecto de las clases sociales y los grupos de interés, y adoptan la racionalidad de los expertos en cuanto a las formas de servicio y eficacia. La evaluación, en opinión de algunos autores, se hará cada vez más fuerte cada vez en el interior de las organizaciones, porque la administración necesita legitimar sus acciones, lo cual puede ser una amenaza para la independencia de los evaluadores. Conceptos como la objetividad, la validez y que el método utilizado sea científico continuarán siendo importantes, aunque también es cierto que el concepto de ciencia está en proceso de redefinición.



4.2. Criterios de actuación profesional

La ética es el conjunto de reglas que lleva al ejercicio correcto de una profesión. Los evaluadores profesionales tienen que evaluar su trabajo para saber en cada momento cuál es la dirección que tienen que seguir con el objetivo de valorar y mejorar la educación.

Actualmente, la mayoría...

... de los evaluadores no ha sido específicamente formadas para serlo. Muchos tienen su formación base en ciencias sociales, pero también en otros campos y ciencias. De este modo, encontramos entre los evaluadores más conocidos a filósofos, psicólogos de la educación, sociólogos, economistas, teóricos del currículum, educadores y gente de otras áreas.

L. Cronbach, en 1964, cuando era presidente de la Asociación de Investigación Educativa Americana (AERA), hizo un llamamiento al desarrollo del conjunto de normas que la evaluación como profesión debía tener. Sin embargo, resultó ser demasiado pronto, aunque ya se esperaba que el número de programas para evaluar creciera considerablemente, y que su influencia en la toma de decisiones aumentara, como de hecho ocurrió. En aquel momento se estaba invirtiendo ya mucho dinero en evaluaciones con informes que no se utilizaban, y los tests estandarizados no se aplicaban siempre bien. Por este motivo, se empezó a generalizar la idea de que era necesario que los consumidores de las evaluaciones tuvieran una guía que les permitiera distinguir entre un buen y un mal trabajo de evaluación (Stufflebeam, 1991).

En 1981, al amparo de un proyecto dirigido por D. Stufflebeam, se constituyó la primera comisión para la redacción de las normas de evaluación, de la cual formaban parte diecisiete miembros que pertenecían a las doce asociaciones profesionales. Más recientemente, en 1994, esta comisión, dirigida por J. Sanders, ha publicado *The program evaluation standards*.

En 1986 se creó la Asociación Americana de Evaluación (AEA), organización dedicada al estudio y la práctica de la evaluación, que aunque está representada en la comisión anteriormente mencionada y apoya sus estudios, nunca ha aceptado esta normativa como propia. En 1992 la AEA decidió poner en funcionamiento una comisión dirigida por P. Rossi que llevaría a cabo esta tarea, sin centrarse en la evaluación educativa como lo hacía el proyecto de Stufflebeam.

Tanto Rossi como los otros miembros del grupo manifestaron, desde los principios de su trabajo, que lo primero que tenían que elaborar era una guía de principios general en lugar de una normativa específica. Entendían que éstos eran más adecuados porque su carácter era más conceptual que operativo, y servirían, por lo tanto, de guía a la actuación del evaluador, en lugar de centrarse en si el evaluador debe tomar decisiones concretas en el terreno metodológico o táctico.

En la comisión de la AEA...

... dirigida por Rossi estaban representados académicos dedicados a la evaluación, evaluadores de agencias privadas y representantes de la administración gubernamental, y aunque en un principio se incluyeron en ella usuarios de la evaluación, parece que se han ido perdiendo por el camino.

Esta decisión se basaba en tres razones: 

1) Los principios son requisitos necesarios para establecer una normativa.

2) Podían desarrollarse con rapidez.

3) Se evitaba disimular con la comisión de "normas" (*standards*).

Después de una serie de discusiones dentro de la asociación, y de las subsiguientes revisiones por el grupo de trabajo, en enero de 1994 el comité ejecutivo de la AEA aprobó los principios propuestos. Los principios para los cuales se consiguió un acuerdo sustantivo fueron los siguientes:

1) Investigación sistemática: los evaluadores tienen que llevar a cabo un proceso investigador con el objetivo de buscar datos sobre el objeto de la evaluación.

a) Los evaluadores tienen que utilizar las normas técnicas más apropiadas, cuantitativas, o cualitativas, para dirigir su trabajo de manera que la información que se utilice y se presente sea precisa y creíble.

b) Los evaluadores deben explorar con el cliente las cuestiones que son esenciales y que tienen que contestarse, así como los enfoques que se deben utilizar.

c) Los evaluadores, al dar a conocer su trabajo, tienen que informar también sobre los enfoques y métodos utilizados, de tal manera que puedan ser bastante conocidos de cara a su interpretación y crítica. Los evaluadores deben discutir en el contexto apropiado y de forma adecuada los valores, las atribuciones, las teorías, los métodos, los resultados y los análisis que afectan de forma significativa a los resultados obtenidos. Esto se debe aplicar a todos los aspectos de la evaluación, desde la conceptualización inicial hasta el eventual uso de los descubrimientos.

2) Competencia: los evaluadores deben proporcionar una ejecución competente a las audiencias interesadas. Este principio se desarrolla en tres puntos:

a) la formación que los evaluadores deben tener;

b) la actualización profesional y el reciclaje al cual tienen que dedicar tiempo para mejorar sus destrezas y pericia, y ofrecer cada vez un servicio de mayor calidad;

c) la recomendación de que las evaluaciones que se lleven a cabo entren dentro de los límites de su formación.

3) Honradez e integridad: los evaluadores deben garantizar la rectitud y honestidad del proceso evaluador en su conjunto. Este principio se desarrolla en siete puntos, que tratan de la negociación tanto con los clientes como con las audiencias sobre costes, metodología, uso de los datos, etc., responsabilidad cuyo inicio corresponde al evaluador. En estos puntos se especifican con detalle las diferentes situaciones que pueden producirse *a posteriori* de la negociación, y se muestra cómo explicitarlas para actuar con transparencia a lo largo del proceso evaluador.

4) Respeto por la gente: los evaluadores deben respetar la seguridad, dignidad y autoestima de los participantes en el programa, clientes y todo tipo de personas integrantes de las distintas audiencias con las cuales se interaccione durante el proceso. Este principio se desarrolla en cinco puntos, en los cuales se trata sobre los problemas y riesgos profesionales que la evaluación puede ocasionar a quienes afecta, pero intentando maximizar los beneficios resultantes de la evaluación y evitando los daños y prejuicios innecesarios. Todo esto se lleva a cabo sin comprometer, por descontado, el compromiso de integridad que todo evaluador contrae al realizar una evaluación y ponerlo en conocimiento de los afectados cuanto antes mejor.

Los evaluadores...

... están obligados a identificar y respetar las diferencias culturales, religiosas, de incapacidades, de sexo, étnicas y otras, de quienes interaccionan en la evaluación.

5) Responsabilidad con el bienestar público: los evaluadores deben articular y tener en cuenta la pluralidad de intereses y valores que se relacionan con el bienestar público. Este principio se desarrolla en cinco puntos, que incluyen el hecho de que los evaluadores deben tener en cuenta, tanto en su planificación como en sus informes, las opiniones y los intereses de todas las audiencias relacionadas con el objeto que se evalúa. Sin embargo, no hay que fijarse

sólo en los efectos inmediatos, sino en los que también pueden producirse potencialmente a más largo plazo, y es preciso hacer accesibles los informes y los datos de la evaluación a las diferentes audiencias. Esto se debe efectuar de tal modo que las necesidades de los clientes y otras audiencias se equilibren, sobre todo en las evaluaciones financiadas con dinero público, y considerando no sólo las necesidades de las audiencias, sino también el bienestar social en su conjunto.

Resumen

La evaluación, de la misma manera que la investigación, produce un conocimiento que pertenece al dominio público, de modo que no es un estudio de uso privado de la administración. La utilidad de la evaluación de programas se basa en su contribución para mejorar la comprensión de las deliberaciones de las políticas públicas, lo cual permite un debate más informado para la acción política; ésta es una aspiración a largo plazo que todos los evaluadores reconocen. El problema de la utilidad de la evaluación no puede tratarse sin considerar la intención social del programa y la responsabilidad de la evaluación, lo cual al mismo tiempo tiene que juzgarse en relación con las posibilidades políticas y el clima del momento.

Con las palabras siguientes se expresaba E. Chemlinsky, presidente de la Asociación Americana de Evaluación en su reunión de noviembre de 1995 en Vancouver, en la apertura de la conferencia que reunía a todas las asociaciones de evaluación que hay en el mundo. Creemos que el texto que ofrecemos a continuación resume mejor de lo que nosotros podríamos hacer la situación en que hoy día se encuentra la evaluación de programas, y que en este módulo hemos querido tratar y mostrar.

> “Durante su corta historia, los evaluadores han reconocido la importancia de centrar de forma precisa las preguntas que necesitan ser contestadas para cada evaluación. Hemos aprendido que la elección de los métodos depende mucho más del tipo de pregunta que se formula que de valorar las cualidades de unos métodos frente a otros. Poner énfasis en la pregunta nos ha liberado de las cuestiones del método y ha permitido que nos centremos en el pluralismo en que la evaluación tiene que establecerse, y superar lo que se hacía en el pasado, cuando se construían evaluaciones para la gloria particular de determinados métodos. Esto ha creado barreras entre los evaluadores, y entre los métodos mismos, que siempre tendrían que haber estado a nuestra disposición.”

Finalmente, puede decirse que la evaluación de programas se ha utilizado hasta el momento al menos con las siguientes finalidades:

– Para supervisar *(monitoring)* programas en acción.

– Para apoyar el desarrollo de un programa nuevo.

– Para identificar los efectos diferenciales de un programa que actúa en distintos grupos de estudiantes u otros tipos de comunidades.

– Para estimar los efectos y costes de los programas y darlos a conocer a los consumidores.

– Para revisar la relevancia y validez de los principios en que se basan los programas.

Estas finalidades previsiblemente se mantendrán, aunque quizá adopten las características particulares de los contextos culturales y políticos.

Actividades

1. Durante una serie de años, una comunidad autónoma ha financiado un programa para enseñar a adultos a leer y escribir. A las personas que han seguido el programa, se les ha pedido que contesten a un cuestionario antes de empezar el curso y al acabarlo, con el fin de comprobar el nivel de alfabetización que ha conseguido cada uno de ellos. La comunidad autónoma contrata a un evaluador externo para que analice los resultados de los cuestionarios o tests y estudie sus efectos en las vidas de los participantes. Pensad de qué tipo es la evaluación que interesa a las autoridades locales y qué características tiene.

2. Escribid un breve ensayo que trate del informe final presentado a la OCDE sobre la enseñanza de las ciencias en la secundaria obligatoria para el proyecto "Currículum e innovación de la enseñanza de las matemáticas, ciencias y tecnología", consultando sobre todo el informe mismo y la metodología utilizada que se incluye en la siguiente publicación española: María José Sáez y col. (1996). *Currículo e innovación en la enseñanza de las ciencias*. INCE.

3. Identificad las audiencias del programa educativo que se especifica a continuación y priorizadlas en función del objetivo de la evaluación y a lo largo de las tres fases principales de la evaluación del programa. Un departamento ministerial de ámbito nacional decide financiar la evaluación de la formación permanente de los profesores de una cierta región, porque está trabajando sobre las líneas maestras que se deben legislar con la finalidad de contribuir a un proyecto de reforma educativa y curricular. El programa pertenece a la comunidad autónoma donde se ha diseñado la formación permanente de los profesores de los centros que implementan el currículum correspondiente. El programa ha establecido una red de responsables en cada provincia de la comunidad autónoma que se encargan y responsabilizan de las actividades marco que tienen cabida dentro del programa, como por ejemplo cursos de perfeccionamiento con diferentes temáticas; proyectos de innovación en el marco de centro escolar o aula; proyectos de investigación con colaboración entre centros escolares y universidades donde se estudian cuestiones como las necesidades formativas de algún nivel educativo concreto o tipo de aprendizaje de profesores que imparten algún currículum específico, etc. El programa cuenta también con la red de centros de profesores y recursos para los cursos que se imparten como apoyo a los proyectos de innovación, así como con los departamentos universitarios de la zona, que colaboran de forma diferente en las distintas actividades.

Ejercicios de autoevaluación

Desarrollo del tema

1. Desarrollad el tema "La evaluación de programas educativos" siguiendo esta pauta:
a) definid *programa* y *evaluación de programas*;
b) describid las consecuencias de la conceptualización de los programas para su evaluación;
c) enumerad algunas razones por las cuales se evalúan los programas;
d) argumentad la naturaleza política de la evaluación;
e) responded a la siguiente pregunta: ¿qué consecuencias tiene para el trabajo de un evaluador la naturaleza cambiante de los programas?

2. Desarrollad el tema "Modelos y métodos para la evaluación de programas educativos" a partir del siguiente cuestionario:
a) ¿Cuál es la principal crítica al modelo de objetivos de comportamiento?
b) ¿Cuál es la aportación más importante del modelo de Scriven a los modelos de evaluación?
c) ¿Qué elementos nuevos proporciona el modelo de estudio de casos a la evaluación?
d) ¿Cuáles son las funciones que se dan al evaluador en los principales modelos?
e) ¿Cuáles son las principales críticas que diferentes teóricos han hecho al modelo experimentalista de la innovación?
f) ¿Cuáles son las preguntas que guían el diseño de la evaluación?
g) Mencionad criterios para la elaboración de un informe de evaluación.

3. Desarrollad el tema "Evaluación, seguimiento y optimización de programas educativos" siguiendo este cuestionario:
a) ¿Cuál es la importancia de las audiencias en la conceptualización actual de la evaluación?
b) ¿Cuáles son las razones que justifican la difusión de los resultados de la evaluación de los programas?
c) ¿Cuál es el origen y fundamento del concepto de negociación en la evaluación de programas?
d) Describid los puntos más débiles de las evaluaciones externas e internas.
e) ¿Cuáles son las principales características y consecuencias de la autoevaluación de programas?

4. Desarrollad el tema "El evaluador como profesional" a partir del siguiente cuestionario:
a) ¿En qué factores se fundamenta el desarrollo de la autoridad profesional?
b) ¿Qué diferencia principal hay entre los principios elaborados por la AEA y la normativa elaborada por el proyecto de Stufflebeam?
c) Resumid los puntos esenciales de los principios elaborados por la AEA.

Solucionario

Ejercicios de autoevaluación

Desarrollo del tema

1.

a) Un programa educativo es un conjunto de actividades organizadas, a veces sin estar totalmente definido como entidad independiente, encaminado a la consecución de unos objetivos que pretenden resolver un problema o atender unas necesidades de política educativa. Estas actividades tienen un sistema de gestión y financiación (al menos por un tiempo), van dirigidas a un grupo de individuos y despiertan el interés de muchos otros (*stakeholders*). La mayoría de los programas educativos en nuestro país depende de un organismo público, de manera que la evaluación que se efectúa forma parte de una situación pública, lo cual implica que el evaluador deba tener en cuenta los intereses colectivos.

La evaluación de programas es un conjunto de métodos, destrezas y sensibilidades para determinar si las actividades llevadas a cabo en la implementación de un programa son necesarias, si se utilizan, si son las adecuadas para suplir la necesidad identificada, si se proporcionan en los términos en que fueron planificadas y si ayudan a la gente que lo necesita, dentro de un coste razonable y sin provocar efectos no deseados.

b) Algunos evaluadores han considerado el programa como una "caja negra", con unos contenidos que no eran el principal objetivo de la evaluación, sino sus efectos. Esto ha llevado a los evaluadores a no conocer bien el programa y, por lo tanto, a no perseguir en la investigación de la información sus aspectos sustanciales, por lo que será difícil establecer las relaciones entre las actividades y sus efectos. En los últimos años se observa un énfasis creciente por definir los programas, algo a lo que los evaluadores cualitativos han contribuido de manera importante.

Hace algunos años, ciertos autores argumentaron en contra de la "caja negra" como conceptualización para la definición del programa (que fue predominante durante un par de décadas, cuando los enfoques experimentales eran los que se aceptaban mayoritariamente), porque al entender que los programas son entidades complejas producidas por una gran cantidad de factores que interactúan entre ellos, y que producen por lo tanto una importante variedad de resultados, consideraban que determinar las relaciones contingentes entre el programa y sus resultados no es algo sencillo. Estos autores defendían la idea de que la definición tenía que hacerse de acuerdo con la teoría de las ciencias sociales, porque una fuente importante de errores para los evaluadores se debe al uso de una versión estandarizada y simplificada de la teoría de la causalidad. Esto no tiene que hacer pensar que los métodos experimentales o casi experimentales no son los más adecuados para evaluar determinados programas, sino más bien que los resultados deben ser interpretados con más cautela, y que se impone que los evaluadores sean más modestos en sus conclusiones en tanto que estos métodos, y en general todos los que se utilizan, no pueden producir unas relaciones causa-efecto tan precisas. Por lo tanto, el modo en que se define el programa afecta de forma sustancial a la evaluación, y por este motivo encontramos la necesidad de convertirlo en objeto de estudio.

c) La evaluación de programas se ha utilizado hasta ahora con las siguientes finalidades:

- para supervisar (*monitoring*) programas en acción;
- para dar apoyo al desarrollo de un programa nuevo;
- para identificar los efectos diferenciales de un programa que actúa en grupos diferentes de estudiantes u otros tipos de comunidades;
- para valorar los efectos y costes de los programas y darlos a conocer a los consumidores;
- para revisar la relevancia y la validez de los principios en que se basan los diferentes programas.

d) La evaluación, como la investigación, produce un conocimiento que pertenece al dominio público, y tanto la educación como los programas educativos afectan al interés colectivo de la comunidad. De esta manera, la evaluación forma parte de una situación pública; es un asunto de tipo colectivo, que hace que se deba tener en cuenta la comunidad de intereses colectivos pertinentes. De hecho, éste es el problema fundamental del evaluador moderno: su evaluación no tan sólo debe ser veraz y creíble, sino que también debe ser justa, pues la evaluación es una empresa que tiene lugar en un contexto político. C. Weiss, que ha estudiado el uso de la evaluación en departamentos ministeriales, argumenta la naturaleza política de la evaluación de programas basándose en tres factores:

- los programas son creaciones de las decisiones políticas, que se mantienen o no, por presión de determinadas fuerzas políticas;
- la evaluación, dado que debe informar la toma de decisiones, tiene un papel en un escenario claramente político;
- la evaluación trabaja facilitando información sobre la legitimidad de los objetivos de un programa y sus estrategias, y sobre la utilidad de éstas para conseguir reformas.

e) Los programas han sido considerados entidades invariables y su variabilidad como una fuente de error. Por este motivo, algunos evaluadores pedían que los programas tuvieran un

tratamiento y una implementación lo más estandarizados que fuese posible en sus múltiples aplicaciones. Esta recomendación se hacía con la finalidad de mantener el rigor del proceso de investigación. Las evaluaciones llevadas a cabo constatan la tendencia de los programas a cambiar, algunas veces de forma sutil y otras de forma repentina y drástica, a causa de múltiples factores.
Esto hace que el evaluador tenga que estar en contacto permanente con los miembros del programa para saber qué cambios se producen y qué efectos tienen sobre las categorías descriptivas que ha utilizado, porque ya conocemos la importancia que tiene para el proceso evaluador definir el programa de forma detallada y precisa. La definición del programa por parte del evaluador tiene que cumplir dos características, aceptadas hoy día por toda la comunidad evaluadora:

– que sea un modelo dinámico del programa que permita asignar categorías y se pueda conceptualizar en función de sus cambios;
– que no sólo se estudien los resultados, sino el proceso mismo de desarrollo del programa y cómo se generan aquéllos.

2.
a) El problema principal es que se basa en el hecho de que la educación es un conjunto de técnicas, que llevan a una finalidad preespecificada, de la que se deriva que los objetivos y los resultados del programa se definen en forma del comportamiento individual de los estudiantes. Limita la definición de los objetivos a valores mensurables y los utiliza como la fuente principal de criterios, que asume que son comunes o aceptados por todas las audiencias. La relación que establece entre las causas (que incluye en las actividades del programa) y los efectos de trabajar con diseños experimentales o casi experimentales es, según algunos autores que han trabajado el tema, demasiado precisa; estos autores recomiendan más prudencia, ya que consideran la educación como un hecho social de una enorme complejidad.

b) Para M. Scriven, uno de los problemas más importantes es el cariz que el evaluador da a la evaluación, lo cual lo lleva a formular la necesidad de que éste desconozca los objetivos del programa para evitar un determinado cariz en la búsqueda de efectos del programa. Al proponer que el evaluador desconozca los objetivos del programa para que éstos no lo condicionen, se plantea el problema tanto en términos de la independencia del evaluador, como de la objetividad del método que utiliza.
M. Scriven centra el resultado de la evaluación en el "producto" al cual ha dado lugar el programa, que es objetivable, pero en el sector educativo es muy problemático identificarlo. Scriven ha estudiado a fondo el problema de la objetividad en la evaluación, y cree que la noción tradicional de subjetividad y objetividad es cuantitativa y se fundamenta sobre todo en el uso de mediciones que son reproducibles. La conceptualización que establece de este término en su modelo es ya cualitativa y se basa en la independencia de las diferentes influencias que interfieren en el juicio.

c) El valor básico reside en el hecho de que los múltiples intereses que hay en el programa tienen que estar representados explícitamente, porque se considera que la evaluación es un servicio a la comunidad, a la cual se da información sobre las características educativas del programa que se evalúa. Por ello, el objetivo de sus informes es hacerlos accesibles al mayor número de audiencias posible, incluyendo a quienes toman las decisiones, y representando el programa como un proceso tanto desde la perspectiva de los evaluadores como de todos los que participan en el mismo. En las evaluaciones que se llevan a cabo con metodologías de tipo naturalista, las reformas y las innovaciones se deben hacer a partir de la investigación. Y el objetivo de ésta tiene que ser informar de la toma de decisiones para conseguir mejoras sociales, y respetar los derechos y las perspectivas individuales de las personas. Las investigaciones naturalistas tienden a evitar la construcción deliberada del control de las circunstancias sociales como forma de obtener nuevas ideas, y pretenden juzgar las acciones en su contexto. Por esto, su objetivo es conseguir una mejor comprensión del programa y representar el trabajo que se está haciendo.

d) En el enfoque de objetivos de comportamiento, la evaluación se encarga de definir los resultados y objetivos del programa en forma del comportamiento individual de los estudiantes, de tal manera que el trabajo del evaluador consiste en observar hasta qué punto los estudiantes muestran estos comportamientos una vez que el programa se ha llevado a cabo. El evaluador especifica los objetivos de forma mensurable y los utiliza como la fuente principal de criterios para comparar, por medio de diseños experimentales o casi experimentales, lo que el programa pretende alcanzar con lo que ha conseguido realmente. La diferencia entre los objetivos y los resultados es lo que da la medida del éxito del programa.
En el enfoque de análisis de sistemas, el objetivo consiste en medir los resultados del programa gubernamental en términos de la relación coste-beneficio. Los evaluadores preparan diseños experimentales en los cuales basan la relación causa-efecto, y eligen el grupo control al azar. Cualquier grupo de evaluadores que realice el trabajo debe obtener los mismos resultados. La reproductibilidad de los resultados es uno de los criterios de validez en este modelo, en el cual objetividad y fiabilidad son sinónimos.

En el enfoque de toma de decisiones, la evaluación se concibe como el proceso de diseñar, obtener y dar información útil para juzgar entre diferentes decisiones posibles, y se estructura en función de la forma en que éstas se tomarán. El evaluador debe identificar de qué manera se toman y definir los criterios que se utilizan en las situaciones identificadas. Además, no planifica experimentos, sino que utiliza el método de muestreo; para él, es crucial identificar las preguntas clave del programa y centrarse en dar información sobre ellas, con las diferentes soluciones, aunque el que toma las decisiones es quien realmente debe contestarlas.
En el enfoque libre de objetivos, el evaluador tiene que buscar todos los efectos que los programas generan y ser un experto capaz de evitar la interferencia debida tanto a él mismo como a los múltiples factores externos. El trabajo del evaluador, según su máximo exponente M. Scriven, es como el de un detective que tiene que descubrir las claves de algo que ya ha sucedido. Es decir, determinar las conexiones causales en condiciones no experimentales. El evaluador tiene que conocer cuál es la aceptación del programa por parte de los consumidores, sin saber los objetivos de las empresas que elaboran el producto, lo cual permite generar criterios de evaluación de acuerdo con las necesidades de los consumidores, que son su principal audiencia de referencia.
En el enfoque de estudio de caso, el objetivo del evaluador se centra en mejorar la comprensión y el conocimiento del objeto evaluado para las audiencias de la evaluación. La función del evaluador es en parte la de un intermediario entre los diferentes grupos de participantes en el programa, de tal manera que quienes trabajan con este enfoque consideran que el instrumento de la evaluación es el evaluador mismo y no las herramientas con que reúne información.

e) Las ciencias sociales se han encargado del estudio de la mejora de las condiciones de vida y de la comprensión de las tendencias, las circunstancias y las fuerzas que actúan en la sociedad, en las que hay una tradición por la cual el progreso social se ha estudiado de acuerdo con la práctica de experimentos. La experimentación proporciona la base para la planificación del cambio social y son los planificadores en colaboración con los evaluadores quienes diseñan y tienen que analizar las innovaciones. Se aplica la lógica científica al estudio y el desarrollo de las innovaciones, para producir un conocimiento preciso, válido y generalizable al servicio de la experimentación social. En estas "sociedades experimentales" se prueban innovaciones para solucionar los problemas, y se confía en la ciencia y la tecnología para decidir las medidas que se deben tomar, ignorando la existencia de fuerzas que han actuado históricamente.
Campbell criticó esta función de la evaluación, como una ciencia aplicada, argumentando que las medidas cuantitativas no pueden sustituir el conocimiento cualitativo, o cómo se pueden manipular fácilmente los objetivos de un programa convertidos en mediciones, o la negación de que las verdades científicas son el producto colectivo de una comunidad en un tiempo determinado, etc.
Cronbach también criticó el planteamiento experimentalista y, aunque defendía que la evaluación es una ciencia aplicada, incorporó el componente político y pluralista a esta visión. La lógica del experimentalismo sugiere que es responsabilidad del evaluador decidir en qué condiciones, circunstancias y poblaciones el tratamiento utilizado en un programa puede generalizarse, y éste depende de la representatividad que las muestras elegidas tienen para practicar el experimento. Según él, la generalización es una cuestión de validez externa que se debe tanto a criterios metodológicos como políticos. House ha criticado también la evaluación como aplicación de la ciencia, porque para él ésta forma parte de los complejos procedimientos de decisión social y moral, y por ello la objetividad científica no trata con la profundidad necesaria el problema de la interferencia, y rechaza así que la validez pueda tratarse estrictamente como una definición operativa.
Sin embargo, el planteamiento experimentalista ha proporcionado una cultura del escepticismo basada en los valores del empirismo, ante el idealismo y el ejercicio administrativo del poder para manipular la información a su conveniencia.

f) Una vez que se define el objeto de la evaluación, el siguiente paso consiste en cómo estudiarlo, para el cual se elabora el diseño, que debe tener en cuenta las siguientes cuestiones:
– ¿Para quién es la información?
– ¿Qué información es la más conveniente?
– ¿Cómo se recopilará la información?
– ¿Quién va a recopilar la información?
– ¿Cómo se comunicará o publicará la información?

g) Para el evaluador, el informe final es el elemento en que se concreta la transacción que ha establecido con el cliente en primera instancia. El informe de evaluación es en primer lugar un canal de comunicación y la representación que el evaluador hace de los hechos que se han producido. Depende en gran medida de cómo se haya formulado la evaluación; en las recomendaciones, los informes en que se haga más énfasis son los que se encaminan a la optimización del programa, pero el evaluador siempre debe tener presente que es preciso que la evaluación cumpla un servicio tanto para clientes como para participantes. La utilidad de la evaluación se concreta muchas veces en el hecho de que, a lo largo del proceso evaluador y del informe, puedan obtenerse opciones de actuación diferentes que permitan intervenir en el programa y enderezarlo. El informe final es un documento dirigido a distintas audien-

cias. A pesar de ello es único, por lo cual se necesita un lenguaje carente de tecnicismos, pero preciso y riguroso, que permita formarse una idea precisa y detallada de lo que es el programa. Usaremos un estilo literario que consiga interesar al lector, lo cual tiene que ver con el estilo narrativo y con la extensión. El argumento se construye utilizando todos los datos recopilados, pero los que se pueden incluir en el informe son pocos. Se recomienda incluir los demás en un anexo para que puedan ser consultados y las inferencias que se hayan producido puedan analizarse. El informe final suele tener carácter analítico, porque pueden ser más breves, lo cual significa organizar la información en los puntos que se consideran más relevantes para la consistencia del programa. El informe debe incluir la metodología en la cual se expone cómo se ha trabajado y una explicación de la elección de la misma.

3.
a) La audiencia de una evaluación son los grupos de individuos que utilizan sus descubrimientos o resultados para reaccionar. Toda evaluación tiene diferentes audiencias y, en una sociedad de mercado, el término consumidor no es ajeno al concepto de audiencia.
Observando la evolución del concepto de evaluación, se observa que las audiencias tienen en el mismo un papel más importante, hasta convertirse en un elemento de la elaboración del diseño. Según algunos autores, en la formulación de las preguntas clave que guían la investigación de la evaluación deben tenerse en cuenta tanto a los usuarios potenciales como a los financiadores, ya que éstos no siempre tienen que considerarse la primera audiencia. Ha sido el reconocimiento de la evaluación como un acto público lo que ha convertido el establecimiento y la argumentación de los criterios de valor en un tema central de estudio en la actualidad.
En los programas educativos hay múltiples criterios de valor, a menudo contrapuestos, porque la gente no quiere lo mismo que los programas públicos, y difiere tanto en sus valores como en sus intereses. Algunos evaluadores eligen la forma tradicional de medir la puesta en marcha de un programa y sus resultados, porque asumen que los objetivos reflejan los intereses de participantes y *stakeholders* y la existencia de consenso en lo que concierne a los valores del programa. La ausencia de este consenso ha hecho que otros evaluadores hayan tratado el problema dando protagonismo a clientes y participantes, tanto a la hora de definir lo que es importante para la evaluación y solicitarles sus opiniones como a la hora de atenderlos como audiencias, e incluso haciéndolos partícipes de algunas decisiones del estudio. Sin embargo, constatar la existencia de múltiples intereses no soluciona cómo conjugar esta multiplicidad para emitir juicios. Los intereses a los cuales sirve la evaluación, y la forma en que éstos se representan son aspectos esenciales de la tarea evaluadora.

b) En cuanto a la publicidad de los informes, dos preguntas clave surgen fácilmente: ¿cuáles son los derechos que tienen quienes financian una evaluación sobre el informe final? ¿cuáles son las obligaciones de los evaluadores para difundir los resultados? Si se quiere optimizar el funcionamiento del programa, los resultados deben ser conocidos por quienes trabajan en él y por los que se ven afectados por su uso. Cuando nos referimos a un programa educativo, los estudiantes son habitualmente los más afectados, aunque muchas veces a causa de la edad no se consideran una audiencia adecuada, pero sí que pueden serlo sus representantes directos, los padres.
La norma propuesta por Rossi y Raizen es que debe llegar a todos aquellos que en función de representar a los beneficiarios del programa reclaman conocer sus resultados e informes. Si la evaluación está financiada con fondos públicos, es el público el que tiene derecho a conocer sus descubrimientos y resultados. De este modo, la evaluación está tratando con el "derecho a conocer" que los ciudadanos tienen en las sociedades democráticas.

c) La negociación en la evaluación es un concepto que tiene su origen en Cronbach, y que está relacionado con el reconocimiento de la pluralidad de intereses. Según Cronbach, lo que se le estaba pidiendo a la evaluación era una información que sirviera de base para una negociación política, en lugar de una información destinada a indicar una decisión correcta; entendía que las políticas sociales en Estados Unidos se estaban configurando por medio de una comunidad de intereses plural, caracterizada por la negociación y el compromiso con las decisiones resultantes, en la cual los políticos articulaban los valores e intereses del público. Esto llevó a formular la naturaleza política de la evaluación, y a no tratarla como una cuestión estrictamente metodológica. House y Cronbach están de acuerdo en que hasta cierto punto una teoría de la evaluación trata de la interacción política y de la determinación de los hechos de los cuales se deben rendir cuentas. Por ello, House interpreta que la validez de la evaluación no puede separarse de las circunstancias y consideraciones políticas y sociales. House centra su análisis de la validez en el componente de intencionalidad social de la evaluación, y argumenta que desde este punto de vista se discriminan ciertas alternativas y ciertos intereses. Cree que se favorece a los más fuertes, representados por las posiciones gubernamentales, los que operan con una definición individualista de los intereses y una atribución pluralista de la distribución del poder que ignora el distinto carácter que hay en las costumbres, las tradiciones y organizaciones administrativas, sociales y políticas.

d) La evaluación interna es la que llevan a cabo los miembros de una organización que no tienen responsabilidad en el diseño y desarrollo del programa que se quiere evaluar. A diferencia de los evaluadores externos, son independientes de los dos.

Se ha debatido durante mucho tiempo qué tipo de evaluaciones son las mejores para un programa, si las internas o las externas, pero ninguno tiene el monopolio de las ventajas. En función de cada situación concreta, se debe analizar qué son ventajas e inconvenientes de una y otra. Hay que tener en cuenta los siguientes aspectos:
1) Confianza de quienes la solicitan.
2) Objetividad y distanciamiento.
3) Comprensión del programa.
4) Potencial de utilización.
5) Autonomía.

e) Desde hace muchos años, las asociaciones de profesionales (médicos, abogados y educadores) han llevado a cabo evaluaciones sobre la formación de los profesionales del sector. Este enfoque es una forma de autoevaluación practicada por profesionales, y en muchas ocasiones por los mismos miembros de la organización objeto de la evaluación. Al final de los setenta, en Inglaterra, quienes hacían desarrollo curricular y evaluación sugirieron que los programas definidos necesitaban centralizadamente iniciativas locales, como una forma de apoyo. La importancia de este apoyo encontró diferentes formas de expresión que pueden entenderse como estrategias organizativas y políticas de reformas arraigadas en las bases. La autoevaluación de centros escolares e instituciones, y el crecimiento de las actividades de evaluación formativa son también formas de expresión de los valores locales.
En Inglaterra, a mediados de los ochenta, los modelos mixtos de evaluaciones, formativas en el ámbito local, y sumativas en el ámbito nacional, se convirtieron en típicas. La autoevaluación con intenciones formativas se enmarca en ciertas concepciones de la mejora de la enseñanza, y muchas de las discusiones llevan a considerar las clases y a los profesores, para saber cómo ayudarlos a cambiar. La importancia del papel del profesor en las teorías de desarrollo curricular y la evaluación es una característica que se debe en parte al énfasis puesto en la autoevaluación como respuesta a la presión de rendición de cuentas de los años setenta. J. Elliott recomendaba entonces el desarrollo de un sistema profesional democrático de rendición de cuentas, aunque creía que la autoevaluación no era solamente una forma de pasar cuentas, sino un instrumento para que los profesores mejoraran su manera de hacer clase y de conocimiento, con unas nuevas formas de profesionalidad. Elliott ha señalado que el énfasis puesto en la rendición de cuentas de las escuelas sólo ha servido para que el poder que éstas tenían fuera a parar a otras instituciones.

4.
a) La legitimación de la autoridad profesional se basa en que:
– el conocimiento y la competencia profesional se validen para la comunidad;
– el conocimiento tenga una fundamentación científica y académica;
– el juicio y el consejo profesional se orienten hacia valores sociales importantes.

b) En 1981, al amparo de un proyecto dirigido por D. Stufflebeam, se constituyó la primera comisión para la redacción de las normas de evaluación. En 1986 se creó la Asociación Americana (AEA), que aunque está representada en el proyecto anteriormente mencionado, decidió en 1992 crear una comisión que llevara a cabo esta tarea. Todos los miembros del grupo manifestaron desde el comienzo de su trabajo que lo primero que tenían que elaborar era una guía de principios general en lugar de una normativa específica. Entendían que éstos eran más adecuados porque tenían un carácter más conceptual que operativo, y servirían de guía a la actuación del evaluador, en vez de centrarse en cómo éste debe tomar decisiones concretas de tipo metodológico o táctico. Esta decisión se basaba en tres razones:
– los principios son requisitos necesarios para establecer una normativa;
– podían desarrollarse rápidamente;
– se evitaba disimular con la comisión de “normas” (*standards*).

c) En lo que respecta a:
– el tipo de investigación que supone el proceso evaluador;
– la competencia y formación de los evaluadores;
– las garantías sobre el procedimiento utilizado para la evaluación;
– el marco para la relación del evaluador con clientes y otras audiencias;
– la responsabilidad con el bienestar público.

Glosario

audiencia interesada
Aquella formada por quienes sin tener relación directa con el programa evaluado se ven afectados directa o indirectamente por el mismo o tienen intereses no necesariamente materiales en su implementación y desarrollo.

audiencia natural
Aquella formada por quienes están directamente relacionados con el programa que se evalúa, ya sea porque trabajan en su implementación, o como población a la cual se dirige.

cliente
Individuo, grupo de individuos u organización que encarga la evaluación.

conclusiones
Juicios finales y recomendaciones emitidos sobre el objeto evaluado.

consumidor
La persona afectada por un programa directa o indirectamente, intencionadamente o no.

contexto
Conjunto de factores que delimitan las circunstancias del objeto evaluado.

criterio
Norma según la cual se emite un juicio.

dato
Información reunida a lo largo de una evaluación.

difusión
Comunicación de las actividades de una evaluación para informar de su conocimiento adquirido y sus resultados.

diseño
Plan de acción que lleva a la práctica de una evaluación.

evaluación
Investigación sistemática encaminada a determinar el mérito y valor de un objeto, ya sea un programa, un proyecto o un producto.

evaluador
Aquel que dirige y lleva a cabo una evaluación.

generalización
Extensión de una información obtenida en un programa a otros que se desarrollan o pueden hacerlo en otros lugares y momentos.

grupo control
Grupo de población estructurado por el evaluador que permite establecer la comparación con el grupo que ha recibido el tratamiento, porque él no lo ha recibido.

información cualitativa
Representación de hechos y opiniones presentados de forma narrativa.

información cuantitativa
Hechos y opiniones presentados de forma numérica.

método experimental
Estudios que se llevan a cabo por medio de la comparación de grupos de individuos en los cuales las variables toman diferentes valores.

utilidad
Influencia de la información y los resultados producidos por una evaluación sobre el programa que se está evaluando.

validez
Propiedad de los argumentos deductivos que son lógicamente impecables.

variable
Característica que puede tomar diferentes valores.

Bibliografía

Bibliografía básica

House, E.R. (1994). *Evaluación, ética y poder*. Madrid: Morata.

Norris, N. (1991). *Understanding Educational Evaluation*. Londres: Kogan Page.

Stufflebeam, D.L.; Shinkfield, A.J. (1987). *Evaluación sistemática*. Barcelona: Paidós-MEC.

Weiss, C.H. (1975). *La investigación evaluativa: métodos para determinar la eficiencia de los programas en acción*. México: Trillas.

Bibliografía complementaria

Adelman, C. (1981). *The Self-Evaluating Institution*. Cambridge Press.

Campbell, T.D. (1983). *Reform as Experiments*. Sage.

Campbell, T.D.; Stanley, J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Skokie. IL, Rand McNally.

Cronbach, L.J. y otros (1980). *Toward Reform of Program Evaluation: Aims, Methods and Institutional Arrangements*. San Francisco: Jossey-Bass.

Eisner, E. (1985). *The Art of Evaluation*. The Falmer Press.

Fernández Ballesteros, R. y col. (1995). *Evaluación de programas: una guía práctica en ámbitos sociales, educativos y de salud*. Síntesis.

Gimeno, J. (1982). *La pedagogía por objetivos: la obsesión por la eficiencia*. Madrid: Morata.

Herman, J.L y col. (1988). *Program Evaluation Kit*. Beverly Hills: Sage.

House, E.R. (1993). *Professional Evaluation. Social Impact and Political Consequences*. Newbury Park, Londres y Nueva Delhi: Sage.

Jaeger, R.M. y col. (1988). *Complementary Methods for Research in Education*. Washington: AERA.

Love, A. (1991). *Internal Evaluation*. Beverly Hills: Sage

McLaughlin, M.W.; Philips, D.C. (1991). *Evaluation and Education: a Quarter Century*. NSSE.

New Directions For Evaluation, colección de revistas de la American Evaluation Association, publicadas por Jossey-Bass.

Patton (1982). *Practical Evaluation*. Beverly Hills: Sage.

Posavac, E.J.; Carey, R.G. (1992). *Program Evaluation Methods and Case Studies*. Prentice Hall.

Raizen; Rossi, P.H. (1981). *Program Evaluation in Education: When? How? To What Ends?* Nat. Academy Press.

Rossi, P.H.; Freeman, H.E. (1993). *Evaluation. A Systemic Approach* (5.ª ed.). Newbury Park, Londres y Nueva Delhi: Sage.

Sáez Brezmes, M.J. (1995). *Evaluación de proyectos, programas e instituciones*. ICE de la Universidad de Alcalá de Henares.

Sáez Brezmes, M.J. (ed.) (1995). *Conceptualizando la evaluación en España*. Alcalá de Henares: Universidad de Alcalá de Henares.

Sáez y col. (1992-1993). Serie de artículos publicados en Acción Educativa desde enero de 1992 hasta junio de 1993.

Sáez, M.J.; Carretero, A. (1994). "El estudio de caso en evaluación o la realidad a través de un caleidoscopio". *Rev. Interuniversitaria de Formación del Profesorado* (núm. 20, págs. 163-178).

Sáez Brezmes, M.J. y col. (1996). *Currículo e innovación en la enseñanza de las ciencias*. INCE.

Scriven, M. (1991). *Evaluation: Thesaurus*. Beverly Hills: Sage.

Stake, R. (1998). *Investigación con estudio de casos*. Madrid: Morata.

Stufflebeam, D. (1988). *Normas de evaluación para programas, proyectos y material educativo*. México: Trillas.

Tesch, R. (1990). *Qualitative Research: Analysis Types and Software Tools*. Falmer Press.